

CHALMERS



GÖTEBORGS UNIVERSITET

Var blåser vinden bäst?

Statistiska undersökningar av vindförhållanden
för vindkraft

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Fredrik Elofsson

Hannes Marling

Amir Mustedanagic

Olivia Säbben

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2012

Var blåser vinden bäst?

Statistiska undersökningar av vindförhållanden
för vindkraft

*Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik
vid Chalmers Tekniska Högskola*

Amir Mustedanagic Hannes Marling Olivia Säbben

*Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk fysik vid
Chalmers Tekniska Högskola*

Fredrik Elofsson

Handledare: Igor Rychlik
Examinator: Carl-Henrik Fant

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2012

Sammanfattning

Med hjälp av en stor mängd vindobservationer i tid och rum analyseras vindhastigheters statistiska egenskaper. Det visar sig att weibullfördelningen väl beskriver vinden, både års- och månadsvis. Vidare används variogram för att undersöka vindens korrelation i rummet. En parametrisk variogrammodell som ger en bra beskrivning för närliggande punkter är Whittle. Dessutom skattas medelvärdesfunktioner för vindens tidsserier. Med hjälp av dessa används kvadratisk optimering för att söka platser där vindkraftverk bör placeras för att minimera energiproduktionens varians. Resultatet av optimeringen visar dessvärre på en stor varians, vilket förklaras med att vinden i det undersökta området är korrelerad och fluktuationar kraftigt. Slutligen föreslås och testas två metoder för skattning av parametrar till den årliga weibullfördelningen med en så kort mätserie som möjligt, dels via en minsta kvadratanpassning och dels via bayesiansk inferens, där Markov chain Monte Carlo används för beräkning av parametrarnas väntevärden. Efter en mätserie på knappt 100 dagar erhålls relativt stabila resultat.

Abstract

Using a large data set, the statistical properties in time and space of wind speeds are investigated. The data is shown to fit the Weibull distribution both on a monthly and yearly basis. Furthermore, a variogram is used to study the spatial correlation of wind speeds. A parametric model for the variogram that well describes the data is Whittle. Moreover, mean functions for the wind speed time series are estimated. The functions are subsequently used in a quadratic optimization problem to find the most optimal positions for wind mills in order to produce a stable energy flow. Unfortunately, the energy output shows a high variance, which is explained by the fact that the wind speeds are correlated and fluctuate heavily. Finally, two methods are proposed and verified to estimate the parameters for the yearly Weibull distribution by using an as short as possible series of measurement. One method is based on fitting by means of least square and another method is based on Bayesian approach, where Markov chain Monte Carlo is used to calculate the expected values of the parameters. After roughly 100 days of measurements, relatively stable results are obtained.

Innehåll

1 Inledning	1
1.1 Bakgrund	1
1.2 Syfte	2
1.3 Avgränsningar	2
1.4 Översikt av genomförande	3
2 Teori	4
2.1 Vindkraftverks energiproduktion	4
2.2 Weibullfördelningen	5
2.3 Skattning av parametrar	6
2.3.1 Maximum Likelihood	6
2.3.2 Bayesiansk inferens	6
2.4 Statistiska tester	8
2.5 Korrelation i tiden	9
2.6 Kovariansstruktur i rummet	9
2.7 Kriging	10
3 Genomförande	12
3.1 Beskrivning av datakälla	12
3.1.1 Beskrivning av det valda området	12
3.2 Undersökning av datamängdens egenskaper	14
3.2.1 Tidsberoende	15
3.2.2 Rumsberoende	15
3.2.3 Fördelning	17
3.2.4 Känslighetsanalys för medelvärdet	21
3.3 Modellerings av medelvärdesfunktion	22
3.3.1 Utvärdering av modellens giltighet: årlig medelvind	26
3.3.2 Utvärdering av modellens giltighet: årlig medeleffekt	28
3.4 Modellanpassning för kortast möjliga tidsperiod	30
3.4.1 Skattning via minsta-kvadrat	30
3.4.2 Bayesiansk inferens	32
3.5 Optimering: bästa möjliga placering av vindkraftverk	36
3.5.1 Utvidgning av området	37
4 Diskussion	39
5 Slutsats	41
Referenser	42
A Fördelningar	44
B Dataomvandling och inläsning i MATLAB	45
C Metropolisalgoritmen	45
D Kriging, fortsättning	46
E Urval av MATLAB-program	47
E.1 Uppskatta parametrar - kortast tid	47
E.2 Anpassa modell och verifiera	48
E.3 Fördelningstest	50
E.4 Visualisering av optimering	50
E.5 Optimering	51
E.6 Bayesiansk parameteranpassning	51
E.7 Skattning av variogram	52

Förord

Vi vill börja med att tacka vår handledare Igor Rychlik för hans kunnande och engagemang.

Alla fyra författarna har bidragit i lika stor utsträckning och med likvärdig arbetsinsats till författandet av denna rapport. Det samma gäller för projektdagboken. Alla författarna har gemensamt stått för kreativitet, problemlösning och kontinuerlig granskning. Vidare har Fredrik haft ansvar för rapportens formatering i \LaTeX .

Nedan redogörs för huvudförfattare för respektive avsnitt. Även om bara ett namn står som huvudansvarig vid vissa avsnitt har ändå alla övriga gruppmedlemmar varit delaktiga och påverkat slutresultatet. Vidare har även samtliga medlemmar fört en individuell tidslogg.

Avsnitt	Huvudförfattare
1	Samtliga
2	
2.1	Fredrik
2.2	Hannes
2.3	Amir
2.4	Olivia
2.5	Hannes
2.6	Amir
2.7	Amir
3	
3.1	Fredrik
3.2	Samtliga
3.2.1	Hannes
3.2.2	Olivia
3.2.3	Olivia
3.2.4	Hannes
3.3	Samtliga
3.3.1	Fredrik
3.3.2	Fredrik
3.4	Samtliga
3.4.1	Fredrik
3.4.2	Amir
3.5	Amir
4	Samtliga
5	Samtliga
Appendix	
A	Hannes
B	Fredrik
C	Amir
D	Amir
E	Se respektive fil

1 Inledning

1.1 Bakgrund

Ett alternativ till fossila energikällor är vindkraft. Det är inte hållbart att fortsätta utsläppen av koldioxid genom nyttjandet av fossila bränslen i den grad mänskligheten gör idag [1]. Vindkraft är en förnyelsebar energikälla med marginell miljöpåverkan.

Nyttjandet av vindenergi har ökat kraftigt under de senaste åren och i Sverige ökade det från 0,63 TWh till 6,1 TWh mellan 2003 och 2011 och trenden är positiv [2]. Enligt en opinionsundersökning utförd av SOM-institutet 2010 är hela 74 % av svenskarna positivt inställda till en utökning av vindkraften [3]. Andelen energi levererad från vindkraftverk uppgick till 4,1 % i Sverige år 2011 [2], vilket kan jämföras med Danmark med en andel på hela 24 % samma år [4].

Vindkraftverk kräver ett relativt litet ingrepp i naturen jämfört med exempelvis vattenkraft. Däremot är energiproduktionen från ett vindkraftverk direkt beroende av den momentana vindhastigheten och kan därför inte tillhandahållas på samma sätt som i ett vattenkraftverk, där vattnet kan lagras och användas vid behov.

Energiproduktionen från vindkraftverk är svår att förutsäga, vilket beror på att vindstyrkan varierar i både rum och tid. För att avgöra om en plats lämpar sig för vindkraftverk räcker det med andra ord inte att provmäta några få dagar. Av både praktiska och ekonomiska skäl önskas dock en så kort mättid som möjligt. En viktig fråga man kan ställa sig är därför under hur kort tid som man behöver mäta för att kunna avgöra om en plats lämpar sig för vindkraftverk.

Även om vind kan verka vara en slumpmässig företeelse finns det ett mönster som kan beskrivas med statistiska modeller. Det verkar vara allmänt vedertaget att den årliga vindhastigheten är fördelad enligt weibullfördelningen [5, 6, 7]. Det är därför rimligt att utgå från denna hypotes.

1.2 Syfte

Syftet med detta kandidatarbete är att, utifrån stora mängder vindobservationer, analysera vindfördelningar över tid och rum. Analysen kommer användas för att avgöra hur ett antal vindkraftverk skall placeras ut, för att en så optimal energiproduktion som möjligt ska uppnås.

Det är även av intresse att bestämma den kortaste tidsperiod, under vilken vindhastigheten behöver mätas, för att utifrån uppmätt data kunna göra en bra förutsägelse av den årliga energiproduktionen från ett vindkraftverk.

Studiens övergripande mål är att använda generella statistiska metoder och tillvägagångssätt som medför att metodiken även kan tillämpas på andra liknande områden.

1.3 Avgränsningar

I studien betraktas problemet primärt ur ett matematiskt perspektiv. Det får till exempel följderna att det inte undersöks om det är möjligt att placera ut vindkraftverk på en viss plats.

Med undantag i den bayesianska inferensen behandlas all data över vindobservationer under antagandet att den är tillförlitlig och korrekt, då det är uppenbart att det saknas både verktyg och kompetens för att analysera och verifiera mätdatan som presenteras. Det är inte någon inskränkning, då projektets fokus ligger på den matematiska modellen, snarare än på resultat från faktiska värden.

Känt är att vindkraftverk skuggar varandras vind, vilket leder till att den totala produktionen från flera närliggande kraftverk påverkas av vindriktningen och antalet kraftverk i omgivningen [8]. Detta är en fråga om vindkraftverkens fysik och kommer därför inte beaktas. Vidare tas ingen hänsyn till vindhastighetens riktning eller variation i höjddled.

1.4 Översikt av genomförande

Projektet har genomförts i ett antal steg, som redovisas nedan.

1. Datakällan med vindhastigheter presenteras och undersöks. Bland annat utvärderas dess tids- och rumsupplösning. Detta beskrivs i avsnitt 3.1.
2. För att få en överblick över datamängden studeras statistiska egenskaper så som periodicitet och tids- och rumsberoende. Vidare bestäms datamängdens års- och månadsvisa fördelning. Detta beskrivs i avsnitt 3.2.
3. En modell för vindhastighetens medelvärdesfunktion tas fram. Denna modell verifieras sedan mot empirisk data. Detta beskrivs i avsnitt 3.3.
4. En uppskattning av den kortaste mättiden som krävs för att avgöra vindhastighetens årliga fördelning görs på två sätt i avsnitt 3.4.
 - (a) Med endast en begränsad mängd data görs en minsta-kvadratanpassning till den parametriska modellen för medelvärdesfunktionen beskriven i avsnitt 3.3. Det undersöks hur stor mängd data som krävs för att parametrarna i denna anpassning ska stabiliseras och konvergera mot modellen beräknad i föregående avsnitt. Detta beskrivs i avsnitt 3.4.1.
 - (b) Med bayesiansk inferens uppskattas parametrar för den årliga vindhastighetsfördelningen, utifrån en begränsad datamängd. Även här beräknas mängden data som behövs för att parametrarna ska konvergera mot den framtagna modellen. Detta beskrivs i avsnitt 3.4.2.
5. Optimeringsproblemet, att placera ut ett antal vindkraftverk så att en så optimal energiproduktion som möjligt uppnås, definieras och studeras med hjälp av resultat från tidigare avsnitt. Detta beskrivs i avsnitt 3.5.

2 Teori

Den teoretiska bakgrunden avseende vindkraftverks energiproduktion och den för vind vanligen använda weibullfördelningen presenteras. Dessutom redogörs kortfattat för de metoder som används för skattning av parametrar, statistiska tester samt korrelation i tid och rum. Slutligen presenteras teorin för *kriging*, som är en metod för att stokastiskt interpolera mellan datapunkter.

2.1 Vindkraftverks energiproduktion

Vindkraftverkens energiproduktion är starkt beroende av vindhastigheten. Om man betraktar vind som ett massflöde av luft med densitet ρ och hastighet v , ges den kinetiska energin, vindenergin, för flödet genom en tänkt area A under tiden t av

$$E_{\text{vind}} = \frac{1}{2}mv^2 = \frac{1}{2}(Avt\rho)v^2 = \frac{1}{2}At\rho v^3. \quad (2.1)$$

Effekten blir då

$$p_{\text{vind}} = \frac{E_{\text{vind}}}{t} = \frac{1}{2}A\rho v^3 \implies p_{\text{vind}} \propto v^3. \quad (2.2)$$

All effekt i ekvation (2.2) kan dock omöjligt utvinnas, för då skulle vinden stanna av direkt efter ha passerat vindkraftverkets rotorblad. Det är därför rimligt att tala om en högsta uppnåelig effekt e_{max} som beror på vindkraftverkets mekaniska och aerodynamiska egenskaper. Ett vindkraftverks effekt vid en viss vindhastighet kan då alltså definieras utifrån detta tal.

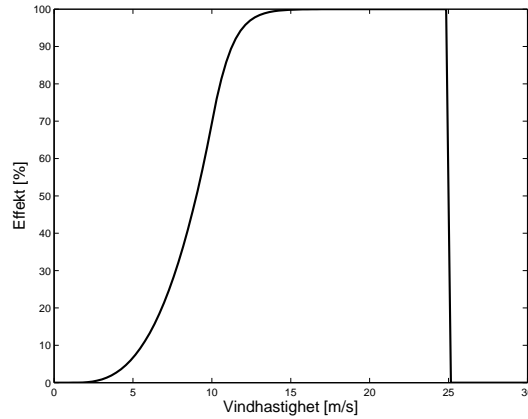
Vidare sker ingen rotation vid alltför låga vindhastigheter på grund av friktion hos rotorbladen och därmed produceras ingen energi. Inte heller vid för höga vindhastigheter produceras någon energi, eftersom vindkraftverk inte klarar av de stora påfrestningar som detta innebär [6]. Med andra ord kan man tala om en *cut-in* och *cut-off* vindhastighet vid vilka den levererade effekten börjar öka respektive minska.

Ett exempel på en funktion som beräknar effekten för ett vindkraftverk, som en andel av den maximala effekten e_{max} , som funktion av vindhastigheten v ges av [9] enligt

$$p(v) = \begin{cases} a_0 \cdot (v - s_{ci})^\beta & \text{om } s_{ci} \leq v \leq s_x, \\ \frac{a_1}{1+e^{-(v-b)^c}} & \text{om } s_x < v \leq s_{co}, \\ 0 & \text{annars,} \end{cases} \quad (2.3)$$

med konstanterna $s_{ci} = 1,2$ m/s; $s_x = 10$ m/s; $s_{co} = 25$ m/s; $\beta = 2,79$; $a_0 = 0,161$; $a_1 = 100$; $b = 9,16$ och $c = 1,05$. Dessa parametrar har skattats med hjälp av tre års högupplöst data från två vindturbiner. Parametern β kan tolkas som en rest av den kubiska termen i (2.2). Funktionen $p(v)$ kan ses i figur 2.1.

Värt att notera är att funktionen (2.3) beräknar en effekt, men en producerad energi beräknas enkelt genom att integrera effekten med avseende på tiden. Därför betraktas begreppen "energi" och "effekt" i detta arbete i princip som synonyma till varandra.



Figur 2.1: Procentsatsen av den maximala effekten e_{\max} ett vindkraftverk levererar som en funktion av vindens hastighet, enligt ekvation (2.3). Vid cirka 5 m/s börjar effekten öka, för att vid ca 10 m/s öka som mest och klinga av mot maximal produktion. Slutligen går effekten mot 0 vid 25 m/s.

2.2 Weibullfördelningen

Som nämnts i bakgrunden är weibullfördelningen en bra beskrivning på hur vindhastigheter är fördelade. Dess täthetsfunktion definieras enligt

$$f_X(x|a, c) = \begin{cases} \frac{c}{a} \left(\frac{x}{a}\right)^{c-1} e^{-(x/a)^c} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2.4)$$

där a kallas skalparameter och c formparameter [10]. En illustration av fördelningen kan ses i figur 2.2.

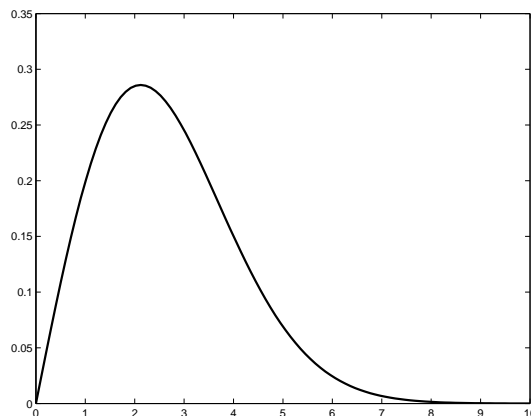
Väntevärdet, μ , av fördelningen ges av

$$\mu = a\Gamma\left(1 + \frac{1}{c}\right). \quad (2.5)$$

Variansen σ^2 ges av

$$\sigma^2 = a^2\Gamma\left(1 + \frac{2}{c}\right) - \mu^2, \quad (2.6)$$

där Γ betecknar gammafunktionen, som definieras enligt $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$.



Figur 2.2: Täthetsfunktion för weibullfördelningen $f_X(x|a, c)$, med parametrar $a = 3$ och $c = 2$.

2.3 Skattning av parametrar

I detta arbete används två olika sätt att skatta parametrar till en fördelningsfunktion, maximum likelihood-metoden och bayesiansk parameteranpassning. I båda dessa metoder antas att $\mathbf{X} = (X_1, \dots, X_n)$ är oberoende observationer av variabeln X . Dessutom antas att X har en parametrisk täthetsfunktion $f(x|\theta)$ där θ är en parametervektor i \mathbb{R}^n . Observera att dessa parametrar är stokastiska, eftersom de är okända. Denna stokastiska vektor, tillhörande θ , betecknas Θ .

2.3.1 Maximum Likelihood

Maximum Likelihood-metoden (ML) maximerar likelihoodfunktionen \mathcal{L} :

$$\mathcal{L}(\theta|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (2.7)$$

Likelihood-funktionen är sannolikheten för det observerade utfallet $\mathbf{x} = (x_1, \dots, x_n)$ som funktion av parametrar [10]. En intuitiv motivering av maximum likelihood-metoden är att den händelse med maximal sannolikhet oftast observeras, då n är stort.

2.3.2 Bayesiansk inferens

I bayesiansk parameteranpassning betraktas parametrarna återigen som en stokastisk vektor Θ . Målet är att utifrån data finna en sannolikhetstäthet för Θ . Denna täthet brukar kallas för posterioritäthet och betecknas som $f_{\Theta}^{post}(\theta|\mathbf{x})$. För att finna $f_{\Theta}^{post}(\theta|\mathbf{x})$ används Bayes sats [10],

$$f_{\Theta}^{post}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)f_{\Theta}^{prior}(\theta)}{f_{\mathbf{X}}(\mathbf{x})}. \quad (2.8)$$

Termen $f_{\Theta}(\theta)^{prior}$ är en vald initial fördelningsfunktion för parametrarna, kallad priorfördelning. Valet av priorfördelningen skiljer sig beroende på vilka parametrar man vill skatta, och det finns inga definitiva regler för hur den ska väljas. Den representerar initial kunskap om parametrarnas fördelning och om sådan saknas väljs den ofta som likformig. Exempelvis kan en lämplig priorfördelning för parametern μ för en normalfördelad variabel vara normalfördelad, på grund av centrala gränsvärdesatsen [10, 11].

Med tillgång till data \mathbf{x} kan termen $f_{\mathbf{x}}(\mathbf{x}|\theta)$ i högerledet beräknas då den är samma som likelihoodfunktionen i (2.7). Det är likelihoodfunktionen som ändrar priorifördelningens utseende till posteriorifördelningens, då nämnaren i Bayes sats endast är en skalfaktor. Nämnaren kan beräknas genom samband mellan betingad och marginell sannolikhetsfördelning enligt

$$f_{\mathbf{x}}(\mathbf{x}) = \int_{\mathbb{R}^n} f_{\mathbf{x},\theta}(\mathbf{x}, \theta) d\theta = \int_{\mathbb{R}^n} f_{\Theta|\mathbf{x}}(\mathbf{x}|\theta) f_{\Theta}(\theta) d\theta. \quad (2.9)$$

Med teorin ovan kan alltså en posteriorifördelning tas fram. För att skatta specifika värden för parametrarna beräknas väntevärdet av Θ genom följande integral

$$\mathbf{E}[\theta_i] = \int_{\mathbb{R}^n} \theta_i f_{\Theta}^{post}(\theta|\mathbf{x}) d\theta \quad (2.10)$$

För approximativ beräkning av (2.10) kan exempelvis *Markov chain Monte Carlo (MCMC)* användas. Då utnyttjas att integralerna är väntevärden som enligt de stora talens lag [10] kan approximeras med

$$\mathbf{E}[\theta_i] = \int_{\mathbb{R}^n} \theta_i f_{\Theta}^{post}(\theta|\mathbf{x}) d\theta = \frac{1}{m} \sum_{j=1}^m \Theta_j, \quad (2.11)$$

för tillräckligt stora m . Θ_j , $j = 1, \dots, m$ är samplade observationer från posteriorifördelningen $f_{\Theta}^{post}(\theta|\mathbf{x})$. Detta är, i stora drag, vad Monte Carlo-metoden i MCMC innebär.

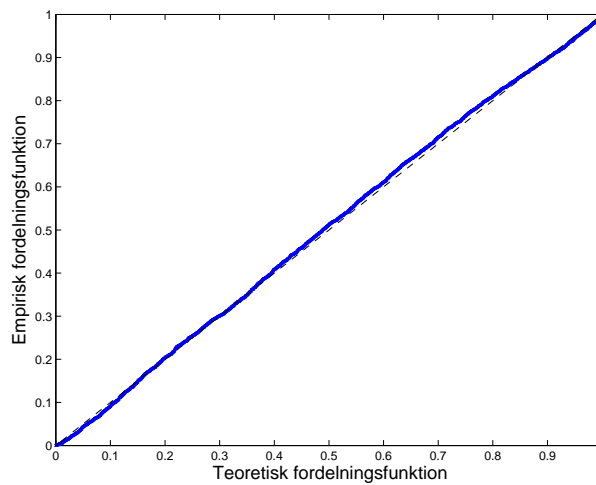
Metropolisalgoritmen, som syftar till Markov chain i MCMC, är en algoritm som ger möjlighet att sampla Θ_j , $j = 1, \dots, m$. Algoritmen är generell och kan sampla observationer från en godtycklig sannolikhetsfördelning. Den utnyttjar egenskaper hos markovkedjor för att slumpa fram observationerna. En utförligare förklaring av hur detta går till ges i appendix C. Fördelen med att använda MCMC-metoden för beräkning av integraler är att den klarar godtycklig dimension och att Metropolisalgoritmen endast kräver en funktion som är proportionell mot sannolikhetstätheten den skall sampla från [12]. Detta innebär att nämnaren i Bayes sats inte behöver beräknas.

2.4 Statistiska tester

Nedan presenteras kort två statistiska tester för att avgöra fördelningar hos stokastiska variabler. Dessutom ges definitionen av p-värde.

PP-plot

En sannolikhet-sannolikhet-plot (probability-probability-plot), eller kort PP-plot, används för att undersöka möjliga fördelningar. Den bygger på att den empiriska fördelningsfunktionen, som i teorin konvergerar mot den riktiga, skattas för en stor datamängd. Den empiriska fördelningsfunktionen plottas därefter mot den antagna fördelningen. Om fördelningarna stämmer överens skall punkterna ligga på en rät linje, se figur 2.3 för ett exempel. Testet ger alltså en grafisk bild av hur bra den antagna fördelningen stämmer överens med den riktiga [10, 11].



Figur 2.3: PP-plot för data mot weibullfördelningen. I detta fall verkar weibullhypotesen högst rimlig då datamängden bildar en rät linje i diagrammet.

Kolmogorov-Smirnov test

Ett kvantitativt mått på hur väl data stämmer överens med en fördelning ges av ett Kolmogorov-Smirnov test (KS-test). Testet ger den största avvikelser i PP-plotten och definieras enligt [13] som

$$\max_i (|F(x_i) - G(x_i)|), \quad (2.12)$$

där $F(x)$ är den empiriska fördelningsfunktionen och $G(x)$ är fördelningsfunktionen för den fördelning som data, x_i , testas mot.

P-värde

Detta mått används när man testar hypoteser. P-värdet är den minsta signifikansnivån α för vilken den påstådda hypotesen, *null-hypotesen*, kan förkastas. Ju mindre det är, desto större är motsägelsen mot null-hypotesen. För tillräckligt små p-värden (p-värde $\leq \alpha$) bör alltså null-hypotesen förkastas [14, 11].

2.5 Korrelation i tiden

Autokorrelation

Autokorrelation är ett mått på den linjära korrelationen i tiden hos en stokastisk process och definieras enligt [15] som

$$r_X(t_1, t_2) = \frac{\mathbf{E}[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]}{\sigma_{t_1}\sigma_{t_2}}, \quad (2.13)$$

där $X(t)$ är en stokastisk process med tillhörande medelvärdesfunktion $\mu_X(t)$, σ_t är variansen vid tidpunkten t och t_1 respektive t_2 är tidpunkter.

Autokorrelationsfunktionen kan till exempel användas för att undersöka hur vinden på en viss plats vid en viss tid beror på hur det tidigare blåst på platsen.

2.6 Kovariansstruktur i rummet

Kovarians i rummet är ett mått på hur två stokastiska variabler, $X_{\mathbf{p}_i}, X_{\mathbf{p}_j}$ samvarierar, då de ligger på olika positioner \mathbf{p}_i och \mathbf{p}_j . Att kovariansen mellan två platser är ett stort positivt tal innebär ofta att det de båda platserna i medeltal avviker åt samma håll i förhållande till sitt medelvärde. Ett stort negativt tal innebär däremot att de i medeltal avviker åt *olika* håll från sitt medelvärde.

Ett vanligt sätt att studera kovarians i rummet är via en kovariansfunktion, C , enligt

$$C(\mathbf{p}_i, \mathbf{p}_j) = \text{Cov}(\mathbf{p}_i, \mathbf{p}_j) = \mathbf{E}[(X_{\mathbf{p}_i} - \mu_{\mathbf{p}_i})(X_{\mathbf{p}_j} - \mu_{\mathbf{p}_j})], \quad (2.14)$$

där $\mu_{\mathbf{p}_i}, \mu_{\mathbf{p}_j}$ är processernas respektive väntevärdesfunktioner. Denna funktion tar alltså in två positioner, $\mathbf{p}_i, \mathbf{p}_j$, och ger tillbaka kovariansen mellan dem.

Dimensionen av denna funktion varierar, beroende på representationen av positionernas koordinater. Antar man att det relevanta området kan representeras av ett tvådimensionellt kartesiskt koordinatsystem blir C fyrdimensionell. Detta motsvarar alltså en representation som ignorerar jordens krökning.

Ett indirekt sätt att framställa kovariansstrukturen är via ett **variogram**, som definieras av [16] som

$$2\gamma(\mathbf{p}_i, \mathbf{p}_j) = \text{Var}(X_{\mathbf{p}_i} - X_{\mathbf{p}_j}) = \text{Var}(X_{\mathbf{p}_i}) + \text{Var}(X_{\mathbf{p}_j}) - 2\text{Cov}(X_{\mathbf{p}_i}, X_{\mathbf{p}_j}). \quad (2.15)$$

Där 2γ betecknar ett variogram. Anledningen till tvåan i definitionen är att endast γ brukar benämnas semivariogram.

Funktionen i ekvation (2.15) uttrycker alltså variansen av differensen mellan $X_{\mathbf{p}_i}$ och $X_{\mathbf{p}_j}$. Även i detta fall beror dimensionen av γ på representationen av \mathbf{p} .

För att lättare analysera ett semivariogram görs antaganden som resulterar i en lägre dimension. Ett vanligt antagande är att semivariogrammet är *homogent*. Ett homogent semivariogram innebär att funktionens värden är invariant under translation i rummet, det vill säga att funktionens värde endast beror av riktningen och avståndet mellan de två positionerna, och inte specifikt av dess koordinater [17].

I fallet då jorden antas vara platt är γ en funktion av Ortsvektorn mellan de båda punkterna, alltså en funktion av två variabler. I ett homogent semivariogram i \mathbb{R}^2 är alltså kovariansen mellan två stokastiska variabler $X_{\mathbf{p}_i}, X_{\mathbf{p}_j}$ med de kartesiska koordinaterna $(1, 1)$ och $(1, 2)$ samma som mellan variablerna med koordinaterna $(2, 0)$ och $(2, 1)$, eftersom de har samma Ortsvektor, $(0, 1)$.

Vissa semivariogram har även egenskapen att de endast är en funktion av avståndet, h , mellan punkterna. Ett sådant semivariogram kallas *isotrop* [17]. Det kommer att undersökas huruvida antagandet om homogenitet och isotropi hos semivariogrammet för vindstyrkan är rimligt.

Då semivariogram används i tillämpningar behövs en modell som beskriver $\gamma(h)$. Det empiriska semivariogrammet approximeras då med en parametrisk modell. Det finns flera välanvända parametriska modeller för semivariogram, däribland exponentiell och Whittle [16]. Båda dessa är specialfall av Matérn-modellen, som har den analytiska formen

$$\gamma(h) = \begin{cases} c_0 + c_1 \left[1 - \frac{\Gamma(\nu)}{2^{\nu-1}} \left(\frac{h}{a}\right)^\nu K_\nu\left(\frac{h}{a}\right) \right] & h > 0 \\ 0 & \text{annars} \end{cases}, \quad (2.16)$$

där h är avståndet mellan de två punkterna, a en parameter beroende av h , c_0 och c_1 är konstanter, ν en parameter, K_ν en modifierad Besselfunktion av ordning ν och Γ är Gammafunktionen [18]. För den exponentiella modellen är $\nu = 0,5$, och för whittlemodellen är $\nu = 1$.

2.7 Kriging

Som ett teoretiskt underlag för en utvidgning av optimeringsproblemet definierat i avsnitt 3.5 presenteras här *kriging*. Krigingmetoden kommer i denna rapport endast diskuteras ur ett teoretiskt perspektiv och inga faktiska beräkningar kommer att genomföras.

Kriging är en grupp av stokastiska interpolationsmetoder. Det finns olika typer av kriging, och nedan presenteras teorin för en av dem, "vanlig kriging", *ordinary kriging*. Teorin är i stora drag samma som den angiven i [19].

Antag att man har tillgång till observationer av variabeln X i punkterna \mathbf{p}_i . Observationerna är då ett diskret stokastiskt fält, vilket innebär en finit samling av stokastiska processer som representerar samtliga platsers vindobservationer. Det stokastiska fält som beskriver vinden är däremot kontinuerligt och betecknas $X(\mathbf{p})$. För att approximera fältet $X(\mathbf{p})$ på de positioner där observationer inte är tillgängliga används kriging.

Vanlig kriging utgår från en modell för det stokastiska fältet enligt

$$X(\mathbf{p}) = m + \epsilon(p), \quad (2.17)$$

där m är den konstanta, och ϵ är den stokastiska delen, av fältet. För att skatta realisationen $\hat{X}(\mathbf{p}_0)$ för en given plats \mathbf{p}_0 antas att skattningen kan skrivas som en linjärkombination av de övriga observerade processerna i fältet, det vill säga

$$\hat{X}(\mathbf{p}_0) = \sum_{i=1}^m \lambda_i X(\mathbf{p}_i). \quad (2.18)$$

Skattningen av $\hat{X}(\mathbf{p}_0)$ överförs på det här sättet till att beräkna λ_i . Observera att det är samma λ_i för alla tider. Att samma λ_i skulle ge en bra skattning för alla tidpunkter är inte särskilt troligt, men detta är heller inte målet med interpolationen, utan målet är istället att finna λ_i , sådan att det i medeltal är en bra skattning. Alltså att

$$\mathbf{E}[\hat{X}(\mathbf{p}_0) - X(\mathbf{p}_i)] = 0, \quad (2.19)$$

Då samtliga platsers medelvärden är m . Samtidigt ska variansen

$$V = \mathbf{E}[(\hat{X}(\mathbf{p}_0) - X(\mathbf{p}_i))^2] \quad (2.20)$$

minimeras. Genom att substituera ekvation (2.18) i (2.19) och i (2.20) samt utnyttja att väntevärdet för alla positioner är detsamma fås följande minimeringsproblem

$$\min \mathbf{E}[(\sum_{i=1}^m \lambda_i X(\mathbf{p}_i) - X(\mathbf{p}_0))^2], \quad (2.21)$$

under bivillkoren att

$$\sum_{i=1}^m \lambda_i = 1. \quad (2.22)$$

Lösningen av detta minimeringsproblem leder till ekvationssystemet nedan [19].

$$\begin{aligned} \sum_{j=1}^m \lambda_j \gamma_{ij} + \mu &= \gamma_{i0} \quad i = 1, \dots, m \\ \sum_{j=1}^m \lambda_j &= 1 \end{aligned} \quad (2.23)$$

En härledning av dessa ekvationer kan ses i appendix D.

Om de parametreiska variogrammodellerna inte visar sig vara tillräckligt bra för stora avstånd kan man tänka sig att endast utnyttja de närmsta punkterna i kriging. Metoden *Moving window kriging* föreslår att man ska välja ett cirkulärt fönster med radie r , som är centrerat i \mathbf{p}_0 och endast utföra kriging i detta fönster [20]. Fönstret rör sig sedan till de platser en skattning önskas. Observera att man ska anpassa ett nytt variogram till varje nytt fönster. Fördelen är alltså att det blir mindre fel i variogramanpassningen, nackdelen är att man har färre punkter. Hur avvägningen däremellan skall utföras i detalj, det vill säga valet av r , diskuteras i [20].

3 Genomförande

Projektet har genomförts stegvis, se avsnitt 1.4, där varje nytt steg bygger på tidigare resultat. I detta avsnitt kommer så väl metod som resultat för respektive steg redovisas. Resultaten diskuteras sedan i avsnitt 4.

3.1 Beskrivning av datakälla

Data över vindhastigheter på olika platser har i detta arbete hämtats från *European Centre for Medium-Range Weather Forecasts* (ECMWF). ECMWF är en internationell organisation med högkvarter i Reading, Storbritannien. Deras främsta mål är enligt hemsidan [21] att, fritt översatt, “utveckla och sköta globala modeller och data-assimileringsystem för dynamik, termodynamik och sammansättning av jordens atmosfär”.

ECMWF har genom åren tagit fram flera så kallade *reanalysis* [22]. Det är enorma datamängder som genererats med hjälp av både vädermodeller och faktiska observationer. Den senaste datamängden kallas ERA-interim. Numera sträcker sig denna datamängd från den 1 januari 1979 och framåt, och uppdateras månadsvis.

I ERA-interim lagras en stor mängd väderobservabler, så som temperatur, vindhastighetens nordliga och östliga komponenter, nederbörd och lufttryck, men även mer komplexa mått, exempelvis “totalt molntäcke”. Dessa observabler anges i regel fyra gånger per dag, vid klockslagen 00, 06, 12 och 18 UTC^a [22].

Var någonstans ERA-interims datapunkter ligger är något mer komplicerat att redogöra för. Först och främst kan man vid nedladdning välja ett godtyckligt område att “saxa ur”^b. Om inget väljs erhålls datapunkter spridda över hela jordklotet. Därefter kan man specificera en rumsupplösning, från 0,75° till 3° longitud/latitudgrader. Detta innebär att det finns en datapunkt per 0,75 respektive 3 longitud-/latitudgrad. Eftersom jorden är rund kommer detta medföra en högre koncentration av punkter kring polerna, samt olika avstånd mellan närliggande punkter beroende på var på jorden man befinner sig. Vidare kan man även välja att få datapunkterna utspridda enligt ett “reducerat gaussiskt grid”. Detta innebär att man undviker en stor ansamling av punkter kring polerna, och att avståndet mellan närliggande punkter i princip är konstant över hela jorden.

ERA-interims data lagras i formatet FM92 GRIB och är tillgänglig för allmänheten på [24]. I appendix B beskrivs omvandling och inläsning till MATLAB av sådana filer. ERA-interim erbjuder vindhastigheten vid 10 meters höjd i både nordlig och västlig riktning [22].

3.1.1 Beskrivning av det valda området

I detta arbete valdes en rumsupplösning på 1,5° för ett område till havs beläget utanför Norge, specificerat av hörnen på “rektangeln” (latitud, longitud): (63, -13), (76, -13), (76, 5) samt (63, 5). I figur 3.1 kan detta område ses. Ett område till havs valdes ur ett förenklande syfte för att undvika den komplicerande effekten av terrängformationer på land.

Områdets storlek är i nord-sydlig led ca 1 445 km, och i öst-västlig led 482 km för de två nordligaste punkterna. För de sydligaste punkterna är avståndet öst-västlig 905 km. Avstånden beräknades med Matlab-funktionen `legs`.

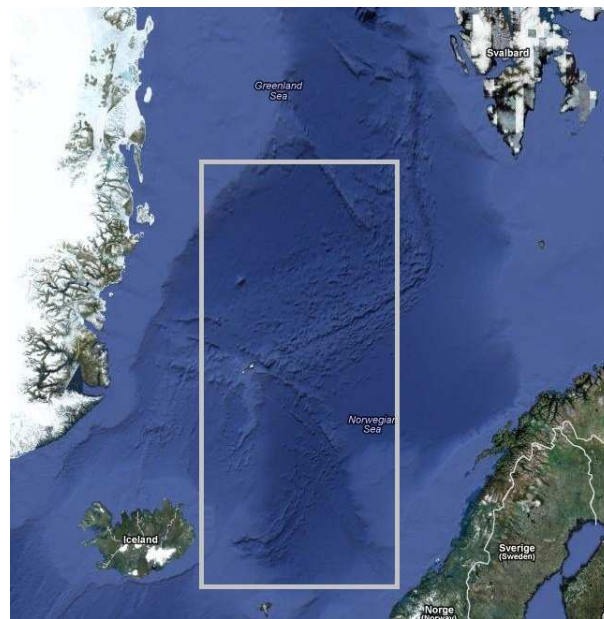
Med rumsupplösningen på en datapunkt per 1,5° longitud/latitud-grad beskrivs alltså detta område av 9×13 punkter, 9 i nord/sydlig och 13 i väst/östlig riktning. För varje punkt

^aUTC är en kompromissförkortning mellan engelskans *Coordinated Universal Time* och franskans *Temps Universel Coordonné*, på svenska “koordinerad universell tid” och är en referenstid världen över. Det innebär i praktiken samma sak som *Greenwich Mean Time* (GMT).[23] Exempelvis anges västeuropeisk tid (Storbritannien, Island, Portugal) som UTC+0. Centraleuropeisk tid (CET) anges som UTC+1 och som UTC+2 vid sommartid. Med andra ord är 12:00 UTC samma sak som 13:00 svensk vintertid.

^bDenna funktionalitet blev dock först tillgänglig i projektets slutskede

laddades vindobservationer från den 1 januari 1979 och ca 31 år framåt ned. Med fyra observationer per dag, blev det ca 45 000 observationer per punkt. Vid inläsningen togs ingen hänsyn till vindens riktning, utan i varje punkt beräknades en vindfart genom att ta beloppet av de båda vindkomponenterna.

Dessa mätningar lagrades sedan i en tredimensionell matris, där raderna representerar latitud, kolonnerna longitud och "djupet" representerar tiden.



Figur 3.1: Karta (© 2011 Google) som visar det utvalda området. Området är specificerat av koordinaterna (latitud, longitud): $(63, -13)$, $(76, -13)$, $(76, 5)$ samt $(63, 5)$.

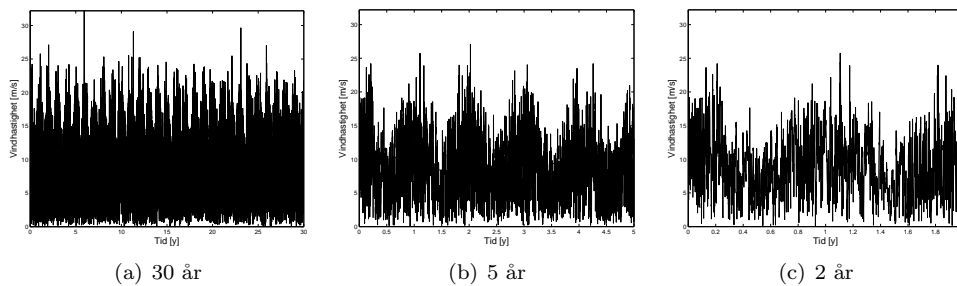
3.2 Undersökning av datamängdens egenskaper

För att få en överblick över datamängden görs inledningsvis några grafiska undersökningar av ett par slumpmässigt valda platser. Därefter studeras datamängdens beroende i tid och rum. Avslutningsvis söks datamängdens årliga och månatliga fördelning.

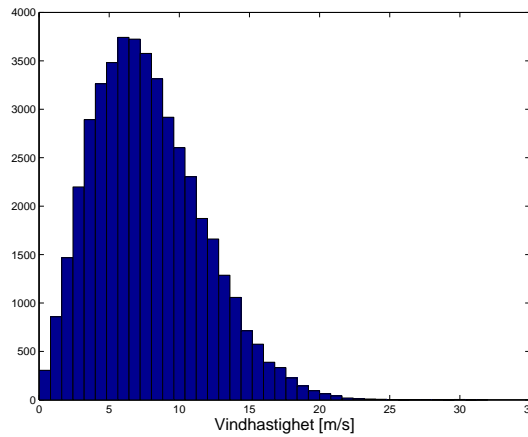
I figur 3.2 ses tidsserier för vindhastighet för några tidsintervall. Som ses har vindhastigheterna ett stokastiskt beteende, men ett årligt periodiskt mönster kan anas.

Ett histogram över alla observationer på en viss plats kan ses i figur 3.3. Medelvärdet för vinden över samtliga mätpunkter på denna plats beräknas till $\mu = 7,7964$ m/s och standardavvikelsen beräknas till $\sigma = 3,879$ m/s.

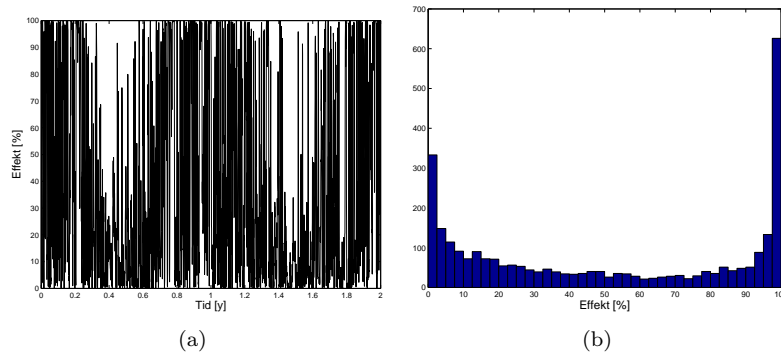
Tidsserie och histogram för den procentuella effekten av den 2-åriga tidsserien av vindhastigheter på samma plats kan ses i figur 3.4. Som ses är effekten starkt fluktuerande. Utseendet på histogrammet i figur 3.4(b) kan förklaras med att effektfunktionen i ekvation (2.3) "klipper" vid låga respektive höga vindhastigheter.



Figur 3.2: Vindhastighet som funktion av tid för en plats i det valda området. Det periodiska beteendet framträder tydligare ur bruset då några få år betraktas.



Figur 3.3: Histogram över en 30-årig tidsserie av vindhastigheter för slumpmässigt vald plats i området.



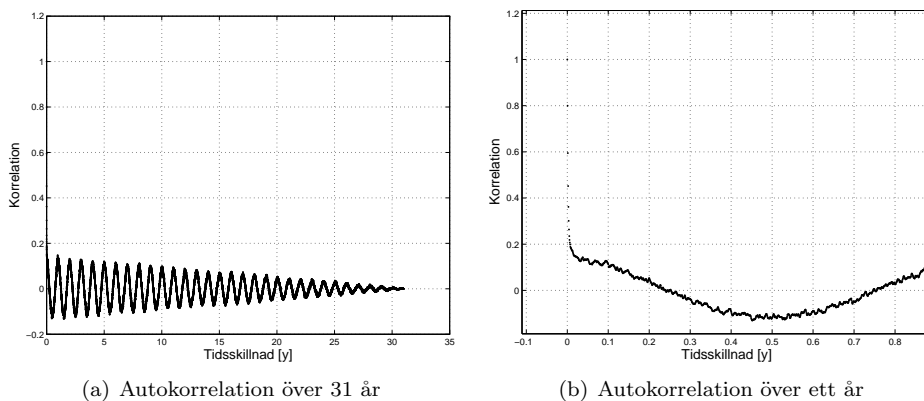
Figur 3.4: Tidsserie och histogram över den procentuella effekten, beräknad enligt ekvation (2.3), av den 2-åriga tidsserien i figur 3.2(c).

3.2.1 Tidsberoende

Vindstyrkans observationer kan betraktas som en stokastisk process i tiden och betecknas $X(t)$. För att undersöka vindens tidsberoende betraktas autokorrelationen av $X(t)$ enligt (2.13).

I förenklande syfte utförs approximationen att korrelationen endast beror av tidsskillnaden, $\tau = t_2 - t_1$, mellan två observationer, $r_X(t_1, t_2) = r_X(t_1 - t_2) = r_X(\tau)$. Processen är då svagt stationär [15]. Med detta antagandet uppskattas korrelationen $r_X(\tau)$ med Matlabs funktion `autocorr`.

I figur 3.5 ses autokorrelationen för en tidsserie i området. I delfigur (a) kan en tydligt årlig periodicitet ses. I delfigur (b) ses en minskning av korrelation från 1 till 0,2 efter cirka 10 datapunkter, motsvarande ungefär 60 timmar. Utifrån denna snabba avklingning görs approximationen att tidsserierna är oberoende i tiden då de spänner över ett avsevärt större tidsintervall. Detta trots att $r_X(\tau) = 0$ inte alltid innebär tidsberoende data.



Figur 3.5: Autokorrelation över 31 respektive ett år för en slumpmässigt vald punkt i området.

3.2.2 Rumsberoende

Vindens rumsberoende undersöks med hjälp av ett semivariogram, γ , som ger ett mått på hur olika punkter i rummet beror av varandra.

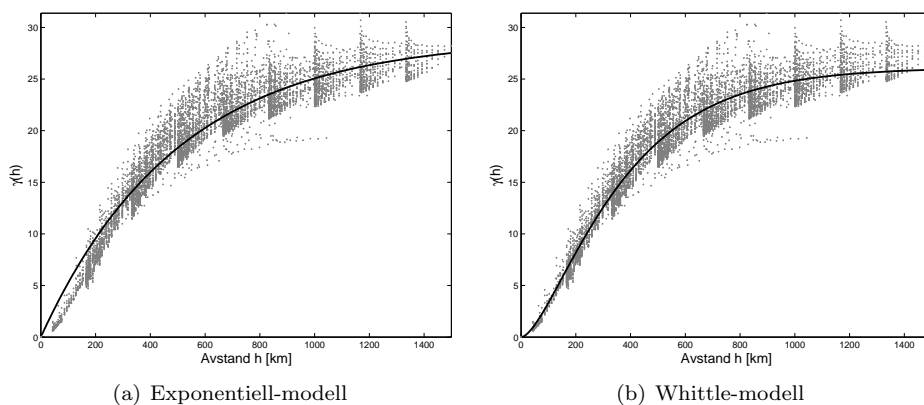
Semivariogrammen beräknas genom att börja i en referenspunkt och beräkna $\gamma(h)$ för avståndet h från referenspunkten till alla andra punkter. I figur 3.6 är $\gamma(h)$ för alla referenspunkter plottade mot alla platser. Det ses att de följer i princip samma mönster och att beroendet

mellan två punkter avtar med ökande avstånd. Det är dock även av betydelse att studera på vilket sätt beroendet avtar.

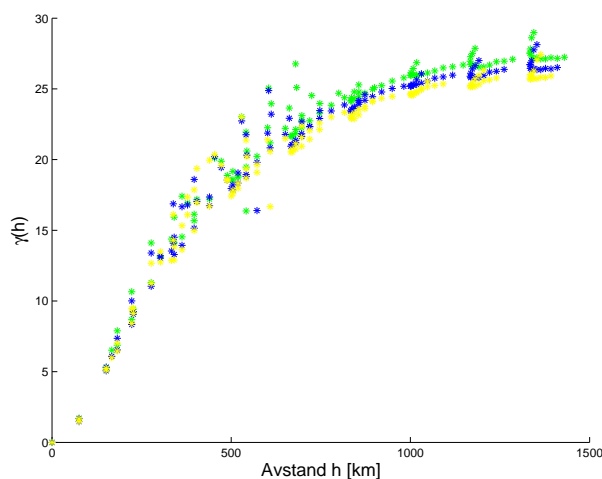
I figur 3.6 ses minstkvadrat-anpassningar av en exponentiell och Whittle-modell mot det skattade variogrammet. Anpassningen för den exponentiella modellen är sämre för kortare avstånd men blir bättre då avståndet ökar, medan motsatsen gäller för whittlemodellen.

Värdena i semivariogrammen uppvisar ett anmärkningsvärt utseende. De är i vissa fall större för kortare avstånd än längre. Detta ses tydligt i figur 3.7 för avstånd $h > 1000$ km som boomerang-former. Det beror på att rumsberoendet mellan vindhastigheter ej är isotropt. Det vill säga att beroendet mellan olika punkter i rummet inte enbart beror på avståndet, utan även på riktningen mellan dem.

Vidare ses i samma figur ses att semivariogrammet för de olika referensplatserna har flyttats ned, vilket indikerar att γ inte är homogent. Med andra ord är semivariogrammet inte invariant under translation i rummet, och beror alltså specifikt på position. Antagandet om homogenitet och isotropi stämmer alltså inte bra i vårt område.



Figur 3.6: Semivariogram över alla platser. Den heldragna svarta linjen är en anpassad modell, se respektive figur. De små grå punkterna är beräknade värden.



Figur 3.7: Semivariogram för tre närliggande referenspunkter.

3.2.3 Fördelning

I detta avsnitt söks en årlig samt månatlig fördelning för vindhastigheterna.

För att kunna beskriva vinddatan och göra förutsägelser om energiproduktion söks en sannolikhetsfördelning för beskrivning av vindhastigheten. Utifrån histogrammet i figur 3.3, väljs sex stycken parametriska täthetsfunktioner ut: weibull, rayleigh, gamma, normal, lognormal och birnbaum-saunders (se appendix A för deras respektive definitioner).

Täthetsfunktionerna (med parametrar skattade med maximum likelihood-metoden) och histogrammet jämförs i figur 3.8. Ju bättre histogrammet följer en fördelnings täthetsfunktion, desto bättre beskriver fördelningen datamängden. Det ses från figuren att weibull, gamma och rayleigh är de fördelningar som bäst beskriver vindobservationerna.

Som en fortsatt undersökning av fördelningsfunktionen för den årliga vindhastigheten görs PP-plottar av de tre ovan utvalda fördelningarna. Detta kan ses i figur 3.9. Ur PP-plotten ses att weibullfördelningen är den fördelning som bäst beskriver vindhastigheten.

För att få ett kvantitativt mått på den maximala avvikelser mellan den empiriska fördelningsfunktionen och den antagna parametriska fördelningen görs ett KS-test med signifikansnivå 0,01. Resultatet av detta test kan ses i tabell 3.1.

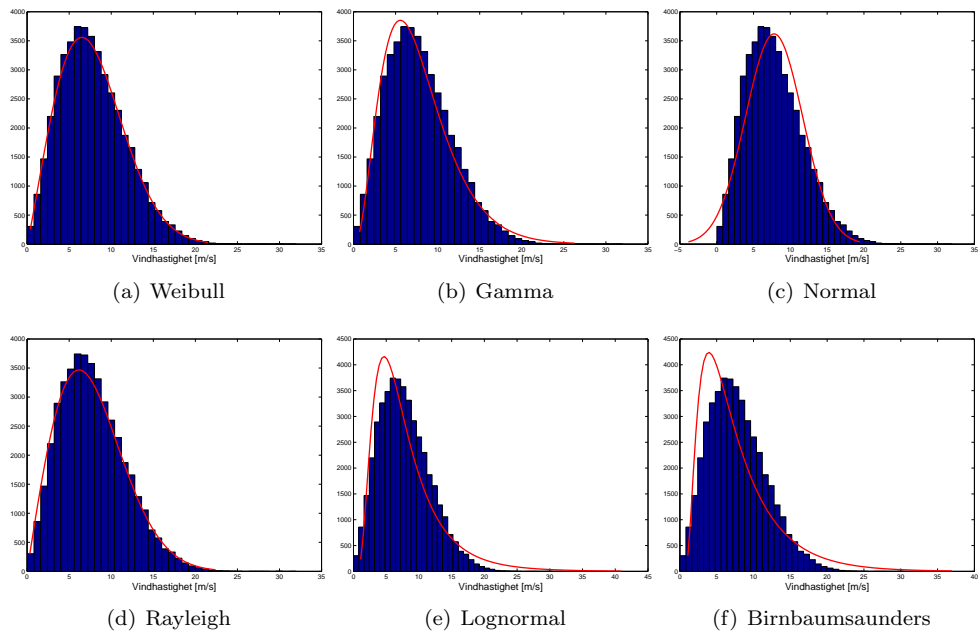
Med stöd av värdena i tabell 3.1 förkastas slutligen gamma och rayleigh. Alltså har alla testade fördelningar, utom weibull, förkastats. Det antas därför att weibullfördelningen är den fördelning som bäst beskriver den årliga vindhastighetsfördelningen, vilket även andra undersökningar har visat [5].

För att undersöka vindhastighetens månatliga fördelning "staplas" observationerna från alla tillgängliga år på varandra. Därefter betraktas ett samplingsintervall på 30 dagar utgående från en dag och 15 dagar framåt och bakåt. I figur 3.12 kan en illustration av detta ses.

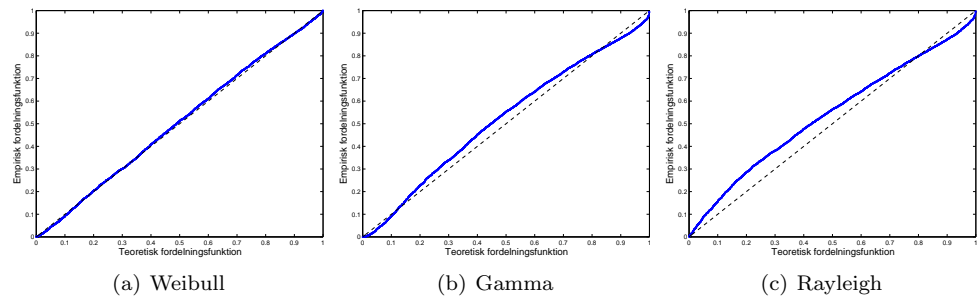
För varje 30-dagarsintervall görs ett KS-test med signifikansnivå 0,01 för att testa weibullhypotesen. Som jämförelse utförs testet även på gamma och rayleigh. Resultatet för dessa fördelningar kan ses i tabell 3.2, som visar hur många gånger respektive hypotes förkastats. Utifrån dessa resultat dras slutsatsen att weibullfördelningen även beskriver hur vindhastigheter är fördelade månadsvis.

Då vindhastigheten kan anses vara weibullfördelad kan fördelningens skal- och formparametrar, a respektive c , uppskattas i en viss punkt över en viss tidsperiod. För varje tidpunkt, t (varje dag), uppskattas dessa parametrar, med maximum likelihood-metoden på samma sätt som KS-testet. Detta görs för alla årets dagar, vilket ger oss 365 skattade parametrar. Hur dessa varierar över året kan ses i figur 3.10. Skalparametern, a , kan ses variera periodiskt med tiden, medan formparametern c verkar vara relativt konstant.

I figur 3.11 ses ett histogram för a och c . Detta fås genom att skatta parametrar med samplingslängd 30 dagar för alla dagar och platser. Pucklarna i delfigur (a) kan ses som ytterligare evidens för skalparameterns periodicitet.



Figur 3.8: Histogram över tidsserie från en slumpmässigt utvald punkt och olika anpassade täthetsfunktioner.



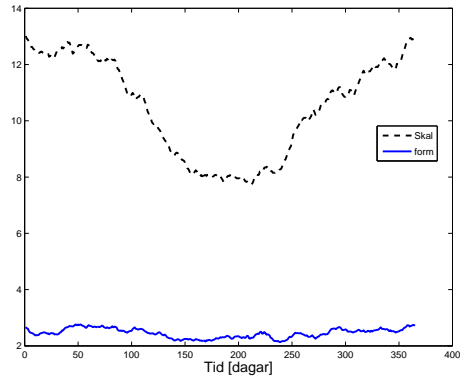
Figur 3.9: PP-plott för weibull, gamma och rayleigh mot samma datamängd som i figur 3.8.

Fördelning	Förkasta	P-värde
Weibull	Nej!	0,0776
Gamma	Ja!	0
Rayleigh	Ja!	0

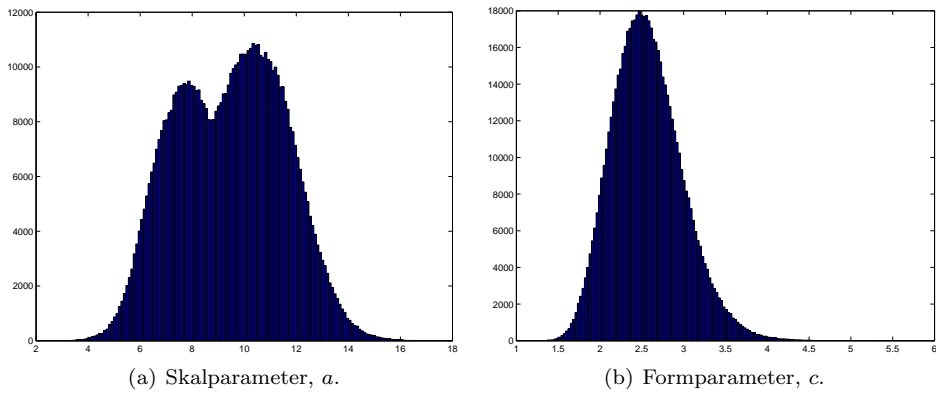
Tabell 3.1: Resultat av KS-test med angiven fördelning som null-hypotes

Fördelning	Antal förkastade hypoteser (av 365)
Weibull	21
Gamma	365
Rayleigh	365

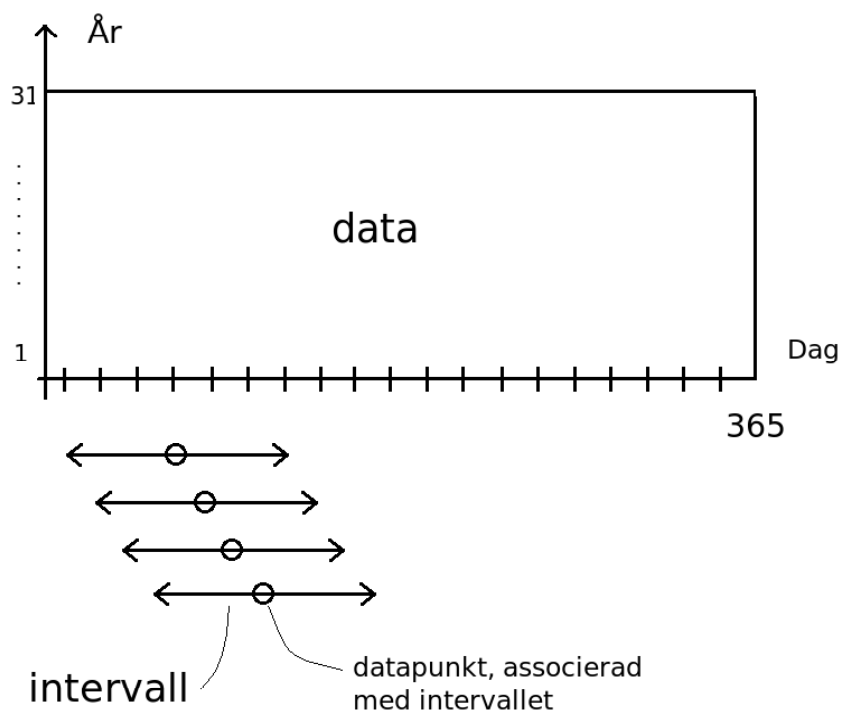
Tabell 3.2: Resultat av KS-test med angiven fördelning som null-hypotes, månadsvis



Figur 3.10: Weibullfördelningens skal- och formparametrar under det “staplade” året för position [1,1].



Figur 3.11: Histogram över den månatliga weibullfördelningens parametrar, för alla punkter i datamängden.



Figur 3.12: En illustration av intervall som successivt stegas fram i tiden.

3.2.4 Känslighetsanalys för medelvärdet

En känslighetsanalys görs för att undersöka hur känsligt medelvärdet för den weibullfördelade vindhastigheten X är, med avseende på formparametern c . Anledningen till detta är att undersöka hur antagandet att formparametern är konstant inverkar på medelvärdet. Medelvärdet för vinden ges av ekvation (2.5) enligt

$$\mathbf{E}[X] = a\Gamma\left(1 + \frac{1}{c}\right). \quad (3.1)$$

Genom att derivera ovanstående ekvation med avseende på c fås

$$\frac{\partial \mathbf{E}[X]}{\partial c} = -\frac{a\psi\left(1 + \frac{1}{c}\right)\Gamma\left(1 + \frac{1}{c}\right)}{c^2}, \quad (3.2)$$

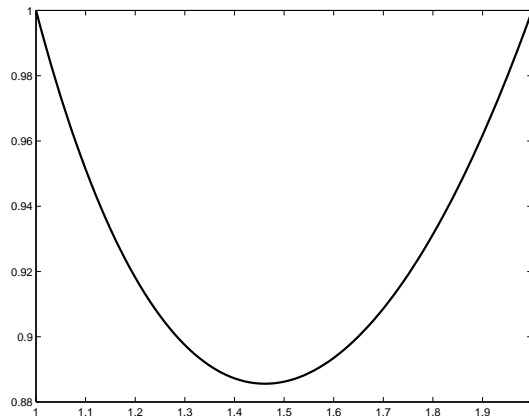
där ψ kallas digammafunktionen och definieras som den logaritmiska derivatan av gammafunktionen [25].

Det största värdet för ekvation (3.2) beräknas nu på intervallet $[\hat{c} - 2c_\sigma, \hat{c} + 2c_\sigma]$, där $\hat{c} = 2,568$ är medelvärdet av samtliga skattade formparametrar enligt figur 3.11 och $c_\sigma = 0,422$ dess standardavvikelse. Som värde på a används medelvärdet från motsvarande skattning enligt figur 3.11.

$$\max_{[\hat{c} - 2c_\sigma, \hat{c} + 2c_\sigma]} \left| \frac{a\psi\left(1 + \frac{1}{c}\right)\Gamma\left(1 + \frac{1}{c}\right)}{c^2} \right| = 0,3141. \quad (3.3)$$

Den ovan gjorda beräkningen visar att beroendet av c är lågt i det aktuella intervallet.

Ytterligare ett argument varför c har en liten inverkan på medelvärdet i ekvation (3.1) kan ses i figur 3.13, som visar gammafunktionen i det relevanta intervallet. Där är gammafunktionen relativt konstant och därmed påverkar c inte medelvärdet i någon större utsträckning.



Figur 3.13: Gammafunktionen, $\Gamma(x)$, i intervallet $[1, 2]$. Notera att x och y-skalan ej är skalenliga.

3.3 Modellering av medelvärdesfunktion

Utifrån tidsserierna av vind i figur 3.2 fluktuerar processen $X(t)$ kring ett periodiskt väntevärde. Vi antar alltså att

$$X(t) = \mu_X(t) + \varepsilon(t), \quad (3.4)$$

där $\varepsilon(t)$ är brus med väntevärde 0, och därmed är

$$\mathbf{E}[X(t)] = \mu_X(t) \quad (3.5)$$

Att uppskatta denna medelvärdesfunktion är viktigt för vidare studier. Exempelvis ger en bra skattning av $\mu_X(t)$ tillgång till vindhastighetens variationer under året och därmed den totala energiproduktionen från ett vindkraftverk. Denna uppskattning är också viktig för att ta reda på hur kort tid man behöver mäta för att tillförlitligt kunna förutsäga energiproduktionen under ett år. Dessutom är den av stor betydelse i optimeringsproblemet som senare skall behandlas, se avsnitt 1.4.

Ett sätt att skatta $\mu_X(t)$ är att använda sig av glidande medelvärde^c. Denna metod har en stor nackdel i att den är helt numerisk och därför erhålls inte någon analytisk koppling mellan vindhastigheten och energiproduktion.

Ett alternativt sätt är att för varje tidpunkt under året beräkna det aritmetiska medelvärdet av samtliga års observationer från denna tidpunkt. Medelvärdet för till exempel den 1 januari beräknas genom att ta medelvärdet av vindhastigheten den 1 januari alla år, enligt: $(1 \text{ jan-79} + 1 \text{ jan-80} + \dots + 1 \text{ jan-10})/31$. Detta konvergerar enligt de stora talens lag mot $\mu_X(t)$ [10]. Nackdelen med att använda denna metod är att då endast 31 observationer från en viss tidpunkt under året är tillgängliga, blir detta medelvärde brusigt. Dessutom saknas återigen en analytisk koppling mellan vind och energi.

Istället utnyttjas att vindhastigheten är månadsvis weibullfördelad. För varje dag under ett år kan alltså parametrar för den månatliga weibullfördelningen skattas. Detta utförs successivt genom att använda data från 15 dagar framåt och 15 dagar bakåt i tiden, för alla dagar på året. Dessa överlappningar kommer medföra en utjämning i tiden av parameterfunktionerna $a(t)$ och $c(t) = c$, liknande det för det glidande medelvärdet. Utifrån en uppskattning av $a(t)$ och ett värde på c (då den kan antas vara konstant enligt känslighetsanalysen i avsnittet tidigare) kan $\mu_X(t)$ beräknas enligt väntevärdet för weibullfördelningen (2.5)

$$\mu_X(t) = a(t)\Gamma\left(1 + \frac{1}{c}\right). \quad (3.6)$$

Som ses i figur 3.14 ger denna modell en bra approximation till det aritmetiska medelvärdet. Vidare ses i figur 3.15 att antagandet att vindhastigheten är månatligen weibullfördelad medför en försumbar skillnad i medelvärde, jämfört med det glidande medelvärdet.

^cDet finns olika metoder för att beräkna det glidande medelvärdet, vi har i detta arbete valt den enklaste formen, nämligen att till varje datapunkt tilldela ett nytt värde somfås genom att beräkna medelvärdet av närliggande punkter. I Matlab beräknas detta med kommandot `smooth`.

Till skillnad från användning av det glidande medelvärdet fås nu en analytisk koppling mellan parametrarnas tidsberoende och energiproduktionen. Den förväntade energiproduktionen vid varje tidpunkt t kan skattas av

$$\mu_p(t) = \mathbf{E}[p(X)](t) = \int_0^\infty p(x)f(x|a(t), c) dx, \quad (3.7)$$

där $p(x)$ är en funktion som omvandlar en vindhastighet till en effektiv effekt, se ekvation (2.3).

Med andra ord kan energiproduktionen och dess medelvärde modelleras på ett effektivt sätt, om man lyckas modellera skalparametern $a(t)$.

En modell för $a(t)$ bör fånga de periodiska huvuddrag som ses i figur 3.10, och samtidigt ha en relativt enkel analytisk formel. I princip skulle en funktion med en mängd av parametrar kunna ansättas, men detta skulle bli opraktiskt vid anpassning till en begränsad mängd data. $a(t)$ anpassas därför till en enkel harmonisk svängning enligt

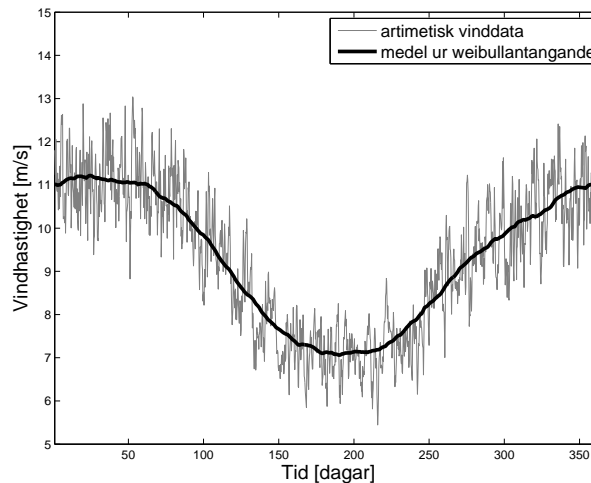
$$a(t) = a_0 + a_1 \cos\left(\frac{2\pi}{T}t + \varphi\right), \quad (3.8)$$

där T är periodtiden på 1 år och φ är en fasförskjutning som motsvarar den tid på året då det blåser som mest, relativt dagen som associeras med $t = 0$.

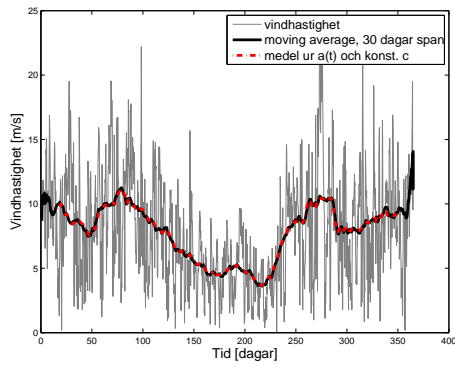
Se figur 3.16 för minstakvadrat-anpassningen av koefficienterna a_0 , a_1 , φ mot skattad månatlig skalparameter. Svängningen följer den skattade skalparametern väl, med undantag av plötsliga pikar.

Eftersom tidsserierna i figur 3.16 startar vid den 1 januari, så motsvarar fasen φ den dag det blåser mest/minst relativt den 1 januari. Allmänt ges denna förskjutning (i dagar) av $365 \cdot \frac{\varphi}{2\pi}$, där ett negativt tecken på φ betyder en "framåtskjutning" och vice versa. Ett värde på $-0,11$ innebär exempelvis en förskjutning framåt i tiden med ca sex dagar och det blåser då som mest den 7 januari.

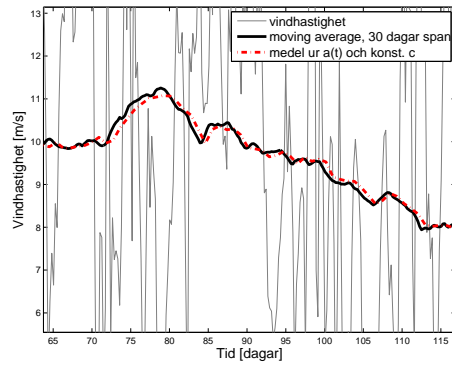
Anpassningen av en periodisk funktion till $a(t)$ är en förenkling, och i nästa avsnitt utvärderas dess giltighet. Dock är det värt att säga att datamängden av skattade skalparametrar direkt kan användas för att finna exempelvis ett års medelvind.



Figur 3.14: Medelvärdet av "staplade" år (grått) och medelvärdet beräknat utifrån weibullparametrar skattade från samma datamängd. Skattningen har skett enligt förfarandet illustrerat i figur 3.12.

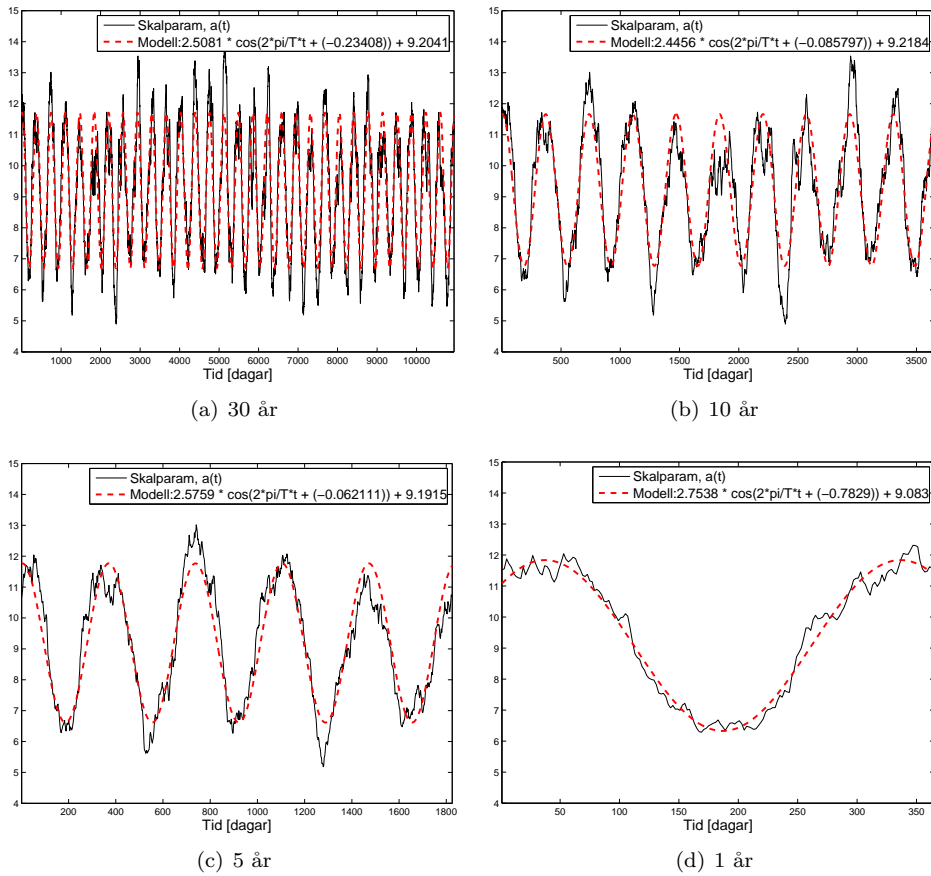


(a)



(b)

Figur 3.15: Vindhastigheter (grått), glidande medelvärde (svart) beräknad med matlabs `smooth` och medelvärde framräknat ur skalparametern $a(t)$ från empirisk data (streckad röd). I båda fallen har ett intervall på 30 dagar använts. Figur b) visar en inzooming av figur a) kring intervallet [60, 115] dagar.



Figur 3.16: Skalparametern $a(t)$ (svart) stegvis framräknad från data ur ett 30 dagars intervall, och anpassad harmonisk svängning (streckad röd) för positionen [7,2] i olika långa intervall, samtliga med sin början den 1 januari 1979. För modellens koefficienter i modellen, se respektive figur.

3.3.1 Utvärdering av modellens giltighet: årlig medelvind

Det *årliga* medelvärdet kan beräknas utifrån det aritmetiska medelvärdet av samtliga observationer. Man kan fråga sig om detta blir samma som det som beräknas från anpassningen till skalparametern i ekvation (3.8). Eftersom en konstant formparameter antagits, kan medelvärdet utifrån (3.8) beräknas enligt (2.5). Medelvärdet under en period $T =$ ett år ges då av

$$\mu_{\text{år}} = \frac{1}{T} \int_0^T a(t) \Gamma\left(1 + \frac{1}{c}\right) dt = a_0 \Gamma\left(1 + \frac{1}{c}\right). \quad (3.9)$$

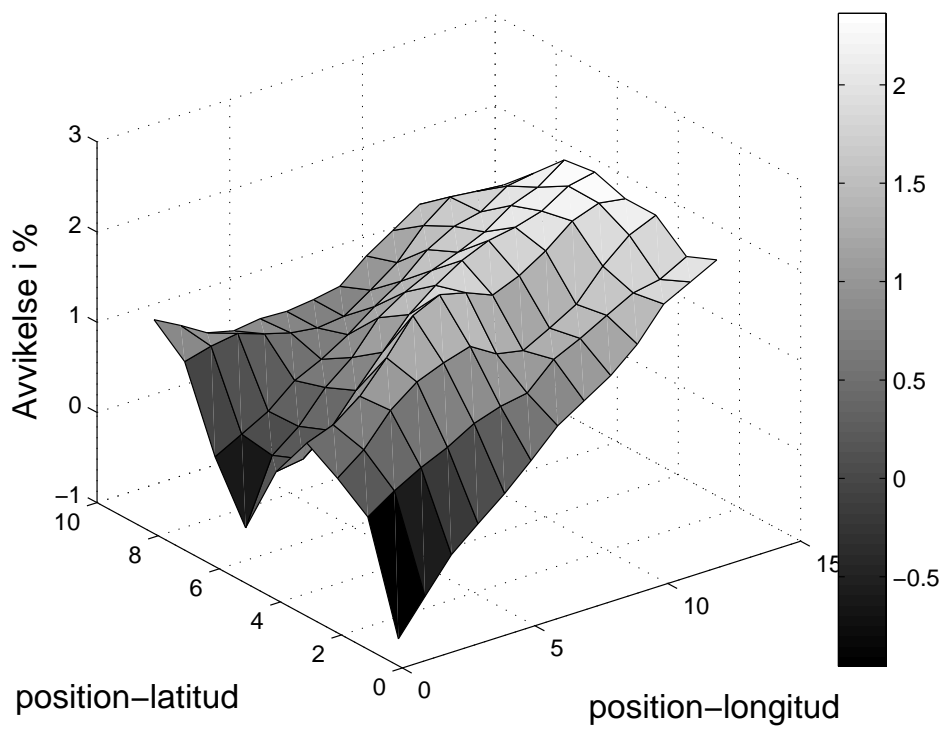
Se tabell 3.3 för en jämförelse mellan det aritmetiska medelvärdet och det genom modellen beräknade.

Se även figur 3.17 för en grafisk jämförelse för alla punkter mellan modellens framräknade medelvind och det aritmetiska medelvärdet. I denna figur ses en ökande avvikelse ju mer åt öster man går, dock som högst ca 2 %.

Alla beräkningar har gjorts med matlab-programmet i avsnitt E.2.

Tabell 3.3: Jämförelse av beräknad medelvind och empirisk medelvind för några olika positioner i datamängden under tidsintervallet 1 januari 1996 - 1 januari 1997. Positionerna anges relativt vårt område, där en ökande longitud/latitud betyder en position mer österut/söderut. Parametern c är skattad från hela året. Den sista kolonnen visar avvikelse från empirisk medel.

“Latitud”	“Longitud”	a_0	c	Medel _{modell}	Medel _{data}	% avvikelse
1	8	10.408	2.3390	8.9935	9.2614	2.8928
1	2	11.100	2.3385	9.6055	9.8061	2.0458
3	8	9.9845	2.2843	8.6223	8.8231	2.2761
3	2	9.7521	1.9942	8.4146	8.6079	2.2459
3	12	9.9666	2.2083	8.6018	8.7736	1.9583
7	2	9.5240	1.9885	8.1593	8.4508	3.4490
7	12	9.9628	2.2131	8.6073	8.7175	1.2641



Figur 3.17: Procentuell avvikelse mellan den empiriskt beräknade medelvinden och den ur modellen beräknade medelvinden för samtliga punkter i datamängden. Tidsintervallet är 1 januari 2001 - 1 januari 2002.

3.3.2 Utvärdering av modellens giltighet: årlig medeleffekt

Det är även intressant att undersöka om modellen förutsäger samma **årliga** energiproduktion som datamängden ger. Enligt ekvation (3.7) kan medelenergiproduktionen under ett år beräknas. Detta jämförs med det aritmetiska medelvärdet av effekter som beräknas av ekvation (2.3) på data.

För att beräkna integralen diskretiseras förloppet i n antal steg. Låt intervallet på 1 år delas upp vid n st tidpunkter enligt $I = [t_1, t_2, \dots, t_i, \dots, t_n]$. Om $n = 365$ svarar alltså t_1 mot 1 januari och t_{365} mot 31 december.

Ekvation (3.7) blir då en summa av delintegraler motsvarande medeleffekt vid tiden t_i

$$\mu_p(t) \approx \frac{1}{n} \sum_{i=1}^n \int_0^{\infty} p(x) f(x|a(t_i), c) dx. \quad (3.10)$$

Ekvation (3.10) kan uppskattas med ett lämpligt val på n . Värdena $a(t_i)$ erhålls genom den harmoniska funktionen i modellanpassningen i avsnitt 3.3.

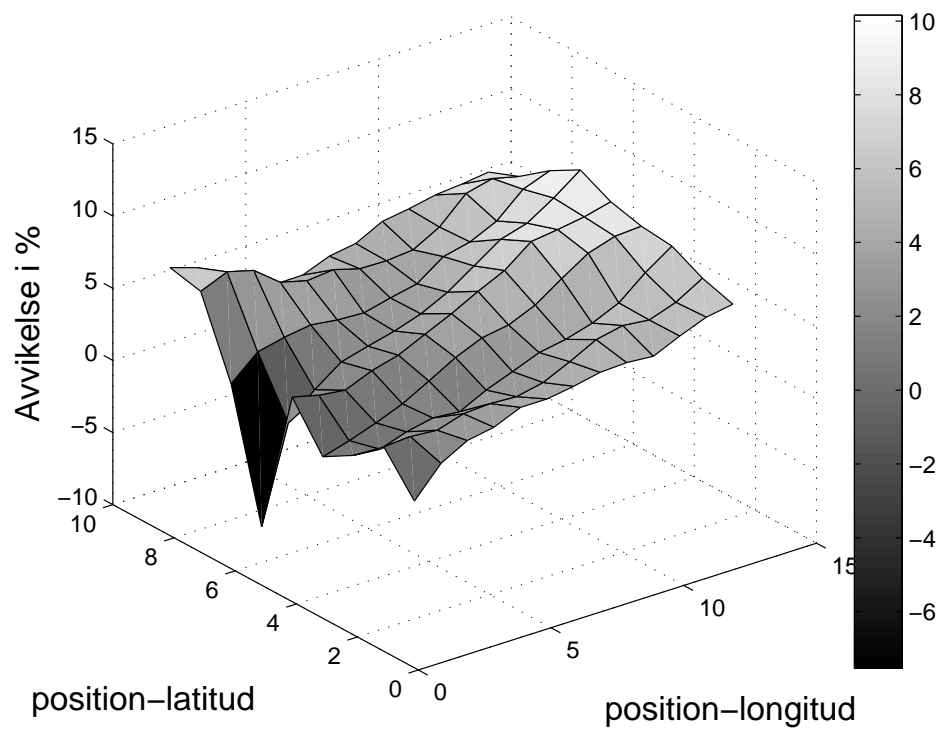
Se tabell 3.4 för en jämförelse mellan aritmetisk medeleffekt och medeleffekt uträknad genom modellen.

Se även figur 3.18 för en grafisk jämförelse mellan alla punkters framräknade modell och empiriskt beräknad effekt. Även här syns en ökande avvikelse för östliga punkter, som högst 10 %.

Även här har alla beräkningar gjorts med matlab-programmet i avsnitt E.2.

Tabell 3.4: Jämförelse av beräknad medeleffekt och empirisk medeleffekt för några olika positioner i datamängden under tidsintervallet 1 januari 1996 - 1 januari 1997. Positionerna anges relativt vårt område, där en ökande longitud/latitud betyder en position mer österut/söderut. Parametern c är skattad från hela året. Den sista kolonnen visar avvikelse från empirisk medeleffekt.

“Latitud”	“Longitud”	a_0	c	Medel _{modell}	Medel-data	% avvikelse
1	8	10.408	2.3390	47.984	52.848	9.2038
1	12	9.8874	2.2412	45.005	47.727	5.7041
3	8	9.9845	2.2843	45.006	48.721	7.6260
3	12	9.9666	2.2083	44.402	47.698	6.9109
7	8	9.6999	2.1923	42.495	46.014	7.6482
7	12	9.9628	2.2131	44.647	47.145	5.2990



Figur 3.18: Procentuell avvikelse mellan den empiriskt beräknade effekten och den ur modellen beräknade effekten för samtliga punkter i datamängden. Tidsintervallet är 1 januari 2001 - 1 januari 2002.

3.4 Modellanpassning för kortast möjliga tidsperiod

I detta avsnitt föreslås två metoder för att bestämma den kortaste mätperiod som krävs för att skatta parametrarna i den årliga weibullfördelningen. Den första är baserad på en minsta-kvadratanpassning och den andra på bayesiansk parameteranpassning.

Det är både av ekonomiska och praktiska skäl önskvärt att under så kort mättid som möjligt uppskatta den förväntade energiproduktionen från ett vindkraftverk på en viss plats. Detta är ekvivalent med att utifrån en mätperiod kortare än ett år skatta parametrar i den årliga weibullfördelningen enligt ekvation

$$\mathbf{E}[p(X)] = \int_0^\infty p(x) f_X(x|a, c) dx, \quad (3.11)$$

där $p(x)$ är funktionen för effekt i ekvation (2.3).

I vårt fall har vi endast tillgång till datamängden från ECMWF, beskrivet i avsnitt 3.1, och inte faktiska mätningar. Då datamängden är mer eller mindre simulerad är den mindre tillförlitlig. För att få en bra skattning av parametrarna i weibull-modellen kan man betrakta datamängden från ECMWF som en utgångspunkt och därefter justera eventuella fel med hjälp av faktiska mätningar. I vårt fall kommer vi dock låta datamängden från ECMWF spela rollen som både expertis och faktiska mätningar.

3.4.1 Skattning via minsta-kvadrat

Som tidigare nämnts gäller att vindobservationer från en punkt i området kan beskrivas av ekvation (3.4). Väntevärdet av denna process är återigen

$$\mathbf{E}[X(t)] = \mu_X(t) \quad (3.12)$$

Som redogjordes i avsnitt 3.3 kan $\mu_X(t)$ beskrivas av funktionen

$$\mu_X(t) = a(t)\Gamma\left(1 + \frac{1}{c}\right), \quad (3.13)$$

där c betraktas som en konstant och där $a(t)$ beskrivs av den harmoniska svängningen i ekvation

$$a(t) = a_0 + a_1 \cos\left(\frac{2\pi}{T}t + \varphi\right). \quad (3.14)$$

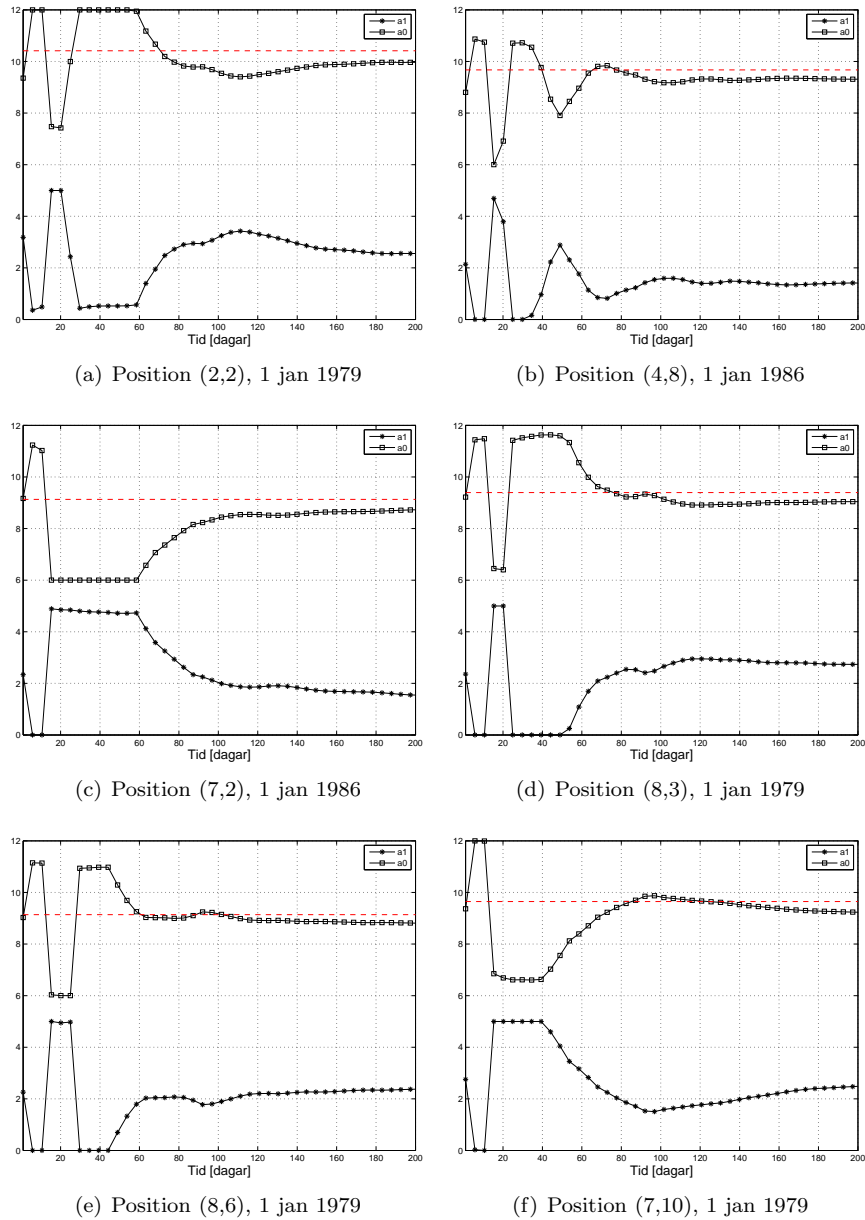
För att skatta $\mu_X(t)$ kan alltså en minstakvadrat-anpassning av koefficienterna a_0 och a_1 i $a(t)$ i ekvation (3.14) utföras. Skillnaden emot föregående avsnitt är att anpassningen nu sker med en begränsad mängd data. Variablerna φ och periodtiden T antas vara kända från tidigare observationer, alltså från ECMWF.

Expertisen utnyttjas även i form av att begränsningar införs på parametrarna a_0 och a_1 . Utifrån figur 3.16 kan man dra slutsatsen att a_0 och a_1 tillhör intervallen $a_0 \in [6, 12]$, $a_1 \in [0, 6]$. Dessa intervall läggs in som villkor i minstakvadrat-anpassningen, se matlab-program i avsnitt E.1.

Se figur 3.19 för den successiva utvecklingen av a_0 och a_1 givet mer och mer mätdata. Det varierar stort hur värdena på a_0 och a_1 utvecklar sig, men efter ca 100 dagar tycks en stabilisering ha skett för båda koefficienterna.

Efter ca 100 dagar, kan man alltså betrakta a_0 som en god uppskattning till den årliga skalparametern. Enligt verifieringen i tabell 3.3 skiljer sig inte $a_0\Gamma\left(1 + \frac{1}{c}\right)$ så mycket från

det årliga aritmetiska medelvärdet och därmed fås en god approximation till den förväntade årliga vindhastigheten. Med skattade a_0 och a_1 kan man nu analogt, enligt avsnitt 3.3.2, beräkna den förväntade årliga medeleffekten.



Figur 3.19: Uppskattade parametrar a_0 och a_1 i ekvation (3.8) som funktion av använd data från ECMWF för olika punkter. Ju mer data som använts, desto stabilare blir skattningarna. Maximum-likelihood-skattningen av weibullfördelningens årliga skalparameter syns streckad i samtliga bilder. Tidsangivelsen i respektive figurtext motsvarar figurens $t = 0$.

3.4.2 Bayesiansk inferens

Metoden beskriven i avsnitt 3.4.1 utgår från att vindstyrka över året uppför sig som den periodiska funktionen ekvation (3.14). Av vinddatan att döma uppför sig vindhastigheten i det valda området på just detta sätt, men då denna data bygger på beräkningar finns inga garantier för att faktiska mätningar skulle uppvisa samma beteende. Ett alternativ till att anta att vindstyrkan uppför sig som i ekvation (3.14) är att istället använda denna funktion som en första utgångspunkt i en bayesiansk inferensmetod. På detta sätt läggs mindre vikt vid att modellen är helt korrekt.

Nedan beskrivs hur Bayes sats, ekvation (2.8), kan tillämpas för att beräkna posteriorifördelningen för skal- och formparametrarna i weibullfördelningen, $f^{post}(a, c|\mathbf{x})$, där $\mathbf{x} = x_1, \dots, x_n$ är en i tid begränsad mätserie. Likelihoodfunktionen är då

$$f(\mathbf{x}|a, c) = \prod_{i=1}^n \frac{c}{a} \left(\frac{x_i}{a}\right)^{c-1} e^{-(x_i/a)^c}, \quad (3.15)$$

Priorifördelningen i ekvation (2.8), kan skrivas via en marginalfördelning som

$$f(a, c) = f_A(a|c)f_C(c). \quad (3.16)$$

Enligt tidigare resonemang varierar c -parametern lite. Därför kan man anta en likformig priorifördelning $f_C(c)$ i intervallet $[c_1, c_2]$ istället för att välja c som en konstant,

$$f_C(c) = \begin{cases} \frac{1}{c_2 - c_1} & \text{om } c_1 \leq c \leq c_2, \\ 0 & \text{annars.} \end{cases} \quad (3.17)$$

För att få en priorifördelning för skalparametern a kan man använda att det samplade årliga medelvärdet M är normalfördelat med medelvärde $a\Gamma(1 + \frac{1}{c})$ och varians σ^2 enligt centrala gränsvärdesatsen. Ett utfall, m , av den stokastiska variabeln M fås genom att för ett år beräkna medelvärdet av samtliga vindobservationer. Observera att enligt problemformuleringen har man ej tillgång till mätningar som sträcker sig över ett helt år. Hur man då uppskattar ett värde på m visas i exemplet nedan.

Fördelningen för skalparametern a , givet formparametern c ges av [26] som

$$f_A(a|c) = \frac{\Gamma(1 + \frac{1}{c})}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(a\Gamma(1 + \frac{1}{c}) - m)^2}{2\sigma^2}\right) \quad (3.18)$$

Koefficienten framför exponentialfunktionen fås från villkoret att (3.18) ska integreras till 1. Genom att bryta ut $\Gamma(1 + \frac{1}{c})$ i exponenten erhålls

$$\frac{\Gamma(1 + \frac{1}{c})}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(a\Gamma(1 + \frac{1}{c}) - m)^2}{2\sigma^2}\right) = \frac{\Gamma(1 + \frac{1}{c})}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(a - m/\Gamma(1 + \frac{1}{c}))^2}{2\frac{\sigma^2}{\Gamma^2(1 + \frac{1}{c})}}\right). \quad (3.19)$$

Variabelsubstitutionen $\sigma^* = \frac{\sigma}{\Gamma(1 + \frac{1}{c})}$ ger nu sannolikhetsfördelningen för en förskjuten normalfördelning

$$f_A(a|c) = \frac{1}{\sqrt{2\pi\sigma^{*2}}} \exp\left(-\frac{(a - m/\Gamma(1 + \frac{1}{c}))^2}{2\sigma^{*2}}\right), \quad (3.20)$$

vilken integreras till 1.

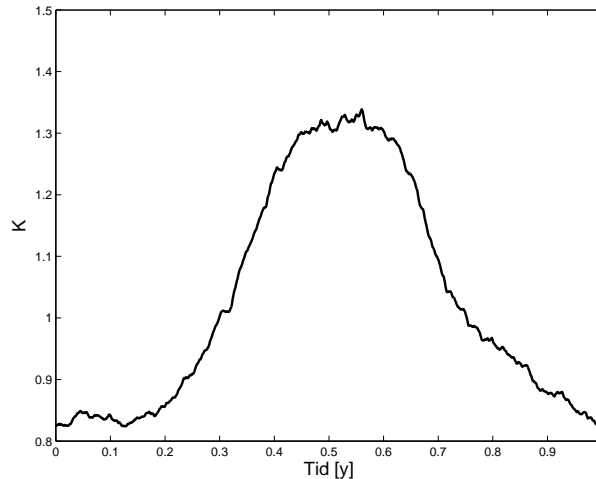
Posteriorifördelningen kan enligt ekvation (2.8) beräknas som

$$f^{post}(a, c|\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}|a, c)}{f(\mathbf{x})} = \frac{f_{\mathbf{x}}(\mathbf{x}|a, c)f_{A,C}(a, c)}{\int_0^\infty \int_0^\infty f_{\mathbf{x}}(\mathbf{x}|a, c)f_{A,C}(a, c) da dc}. \quad (3.21)$$

Som tidigare nämnts i teoriavsnittet 2.3.2 kommer inte nämnaren i ekvationen ovan beräknas då vi använder Markov chain Monte Carlo för att beräkna parametrarnas väntevärden.

Exempel

Innan en posteriorifördelning kan beräknas måste mätningarna justeras för säsongsvariationerna. En metod för att göra det är (som tidigare gjorts i avsnitt 3.4.1) att skatta en medelvärdesfunktion från en befintlig källa, i vårt fall från ECMWF. Därefter beräknas hur mycket den avviker i procent från den befintliga datakällans årliga medelvärde i varje tidssteg. Detta har gjorts för en plats och funktionen kan ses i figur 3.20. Mätningarna multipliceras sedan med denna avvikelse. Det är inte troligt att funktionen i figur 3.20 stämmer för de faktiska mätningarna, men det är inte heller avsikten med den. Avsikten är istället att använda den som utgångspunkt i den bayesianska modellen och låta faktiska mätningar justera eventuella fel.



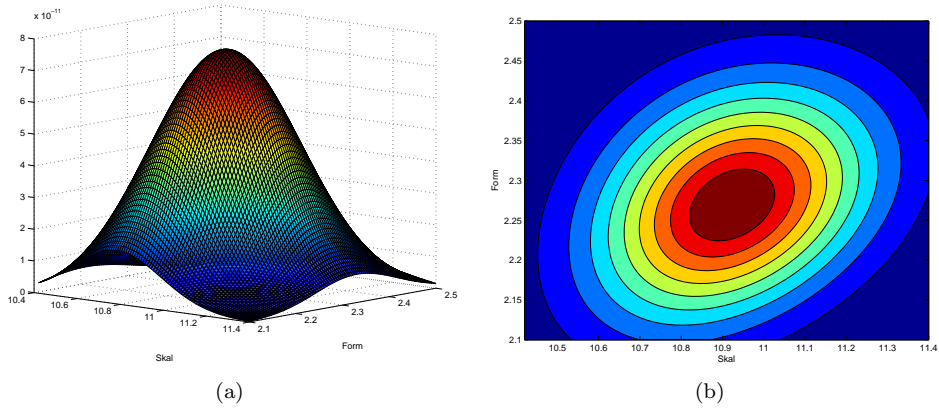
Figur 3.20: Faktorfunktion för justering av säsongsvariationer

I figur 3.21 ses en funktion proportionell mot posteriorifördelningen för en position. Här har observationer från 100 dagar i maj-augusti använts. Mätningarna från dessa hundra dagar multiplicerades med motsvarande element i faktorfunktionen i figur 3.20. Ett värde på m erhöles genom att ta medelvärdet av denna produkt. Variansen väljs till $\sigma^2 = 1$ med motivering att det samplade årliga medelvärdet M inte kommer att ha hög standardavvikelse då antalet mätpunkter är ganska många.

Väntevärden för parametrarna på denna position beräknades med Markov chain Monte Carlo. Skillnaden mellan dessa och de parametrar som skattades med maximum likelihood under hela året blev då 1,80% för a och 0,43% för c . Beräkningen utfördes för olika m i syfte att se hur resultatet påverkas. Resultatet i figur 3.22 visar att metoden är relativt stabil med avseende på denna parameter.

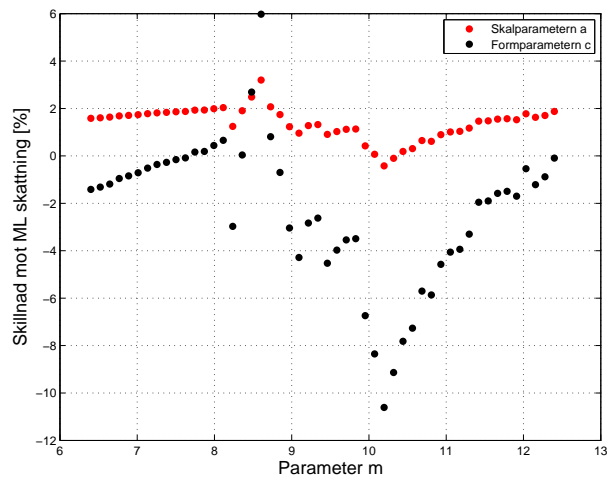
I figur 3.23 ses hur årlig maximum likelihood och bayesiansk skattning skiljer sig beroende på antalet mätdagar. I denna figur är parametern $m = 9,4$ och i figur 23(b) är $m = 7$. Skillnaden mellan dessa är väldigt liten vilket återigen tyder på att skattningen är stabil. Man kan även se att skillnaden är ganska stor i början men minskar och blir stabil efter 80-100 dagar. Skalparametern, som har störst inflytande på medelvärdet, är ganska stabil efter relativt kort tid.

Observera att parameterskillnaden efter 100 dagar i figur 3.23 skiljer sig från de ovan nämnda värdena 1,8 % och 0,43 % trots att det är samma position och parametrar. Detta på grund

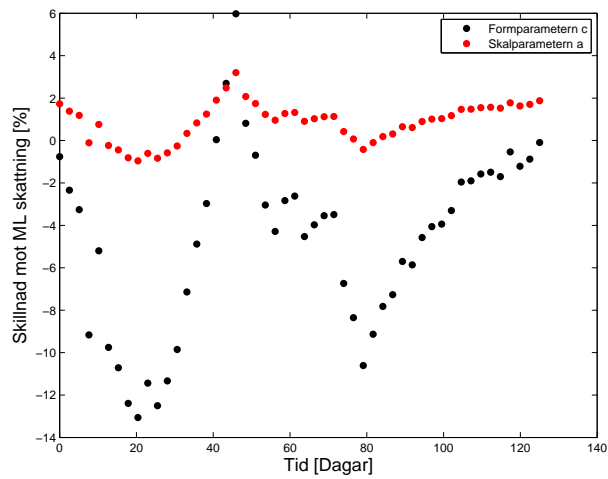


Figur 3.21: En funktion proportionell mot posteriorifördelning

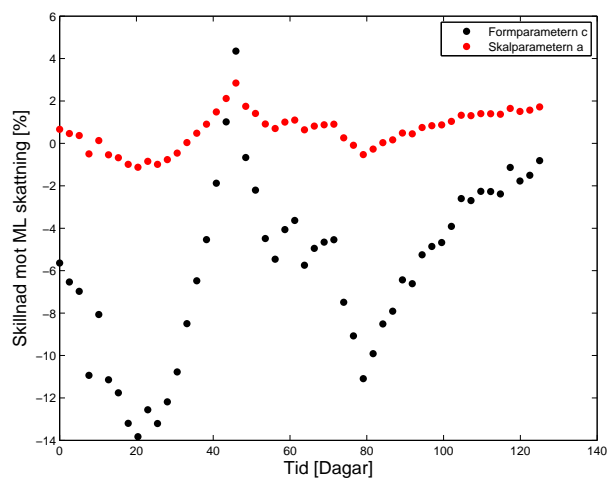
av att skattningen inte är tidshomogen, det vill säga den beror inte bara på antal samplingspunkter utan också mellan vilka datum man skattar. I figuren ses 100 dagar mellan april och juli. Det är generellt bäst att utföra mätningar på sommaren, då variansen hos vinden är som minst.



Figur 3.22: Skillnad mot årliga ML parametrar i procent som funktion av parametern (skal: röda prickar; form: svarta prickar)



(a) Skillnad mot årliga ML parametrar som funktion av antal mätdagar då $m=9.4$



(b) Skillnad mot årliga ML parametrar som funktion av antal mätdagar då $m=7$

Figur 3.23: Posteriorifördelning

3.5 Optimering: bästa möjliga placering av vindkraftverk

Tidigare undersökningar av vindhastigheten visar att den är starkt fluktuerande. Därmed blir även energiproduktionen det, se figur 3.4 på sidan 15. Ur ett energiutvinningsperspektiv, kan det vara önskvärt att skapa en stabil energiproduktion med mindre fluktuationer. Detta skulle exempelvis kunna innebära att man, givet en smart konfiguration av vindkraftverk, kan utnyttja vindkraft som en huvudsaklig energikälla.

Ett sätt att definiera detta problem på är att minimera variansen av den totala energiproduktionen från n stycken vindkraftverk utplacerade på m olika platser. Det är även intressant att införa bivillkoret att väntevärdet av den totala årliga energiproduktionen ska överstiga ett givet värde, K , för att undvika stabila men otroligt låga energiproduktionsnivåer.

Matematiskt kan detta problem härledas på följande sätt. Låt w_i vara andelen, i procent, av de n kraftverk som skall placeras ut på plats i . Låt Y beteckna den stokastiska variabel som motsvarar den totala energin producerad av en given konfiguration av vindkraftverk. Y kan då skrivas som

$$Y = \sum_{i=1}^m w_i e_{\mathbf{p}_i}, \quad (3.22)$$

där den stokastiska variabeln $e_{\mathbf{p}_i}$ motsvarar procentsatsen av den maximala energin e_{\max} som är den maximala effekt som kraftverket i position \mathbf{p}_i , $i = 1, \dots, m$, kan producera. Observera att variabeln Y betecknar energi producerad någon gång under året.

Variansen av Y kan uttryckas som

$$\begin{aligned} \text{Var } Y &= \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 \\ &= \mathbf{E}\left[\left(\sum_{i=1}^m w_i e_{\mathbf{p}_i}\right)^2\right] - \mathbf{E}\left[\left(\sum_{i=1}^m w_i e_{\mathbf{p}_i}\right)\right]^2 \\ &= \mathbf{E}\left[\sum_{i=1}^m \sum_{j=1}^m w_i w_j e_{\mathbf{p}_i} e_{\mathbf{p}_j}\right] - \left(\sum_{i=1}^m w_i \mathbf{E}[e_{\mathbf{p}_i}]\right)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^m w_i w_j \mathbf{E}[e_{\mathbf{p}_i} e_{\mathbf{p}_j}] - \sum_{i=1}^m \sum_{j=1}^m w_i w_j \mathbf{E}[e_{\mathbf{p}_i}] \mathbf{E}[e_{\mathbf{p}_j}] \\ &= \sum_{i=1}^m \sum_{j=1}^m w_i w_j (\mathbf{E}[e_{\mathbf{p}_i} e_{\mathbf{p}_j}] - \mathbf{E}[e_{\mathbf{p}_i}] \mathbf{E}[e_{\mathbf{p}_j}]) \\ &= \sum_{i=1}^m \sum_{j=1}^m w_i w_j \text{Cov}(e_{\mathbf{p}_i}, e_{\mathbf{p}_j}) \\ &= \mathbf{w}^T \mathbf{A} \mathbf{w} \end{aligned} \quad (3.23)$$

Där \mathbf{A} är kovariansmatrisen, $\text{Cov}(e_{\mathbf{p}_i}, e_{\mathbf{p}_j})$, över datamängden. Därmed fås följande kvadratiske minimeringsproblem

$$\min \mathbf{w}^T \mathbf{A} \mathbf{w} \quad (3.24)$$

Under bivillkoren att:

$$\mathbf{w}^T \cdot \boldsymbol{\mu} > K \quad (3.25a)$$

$$\sum_{i=1}^m w_i = 1 \quad (3.25b)$$

$$w_i \geq 0 \quad (3.25c)$$

där μ är en vektor med de aktuella positionernas energimedelvärde tagna över all data och $K > 0$ är en vald konstant som representerar tröskeln för den efterfrågade effekten.

Då observationerna av $e_{\mathbf{p}_i}$ är tidsserier $e_{\mathbf{p}_i}(t)$ kan kovariansmatrisens element a_{ij} skattas [10],

$$a_{ij} = \frac{1}{N-1} \sum_{k=1}^N (e_{\mathbf{p}_i}(t_k) - \mu_{\mathbf{p}_i}(t_k))(e_{\mathbf{p}_j}(t_k) - \mu_{\mathbf{p}_j}(t_k)). \quad (3.26)$$

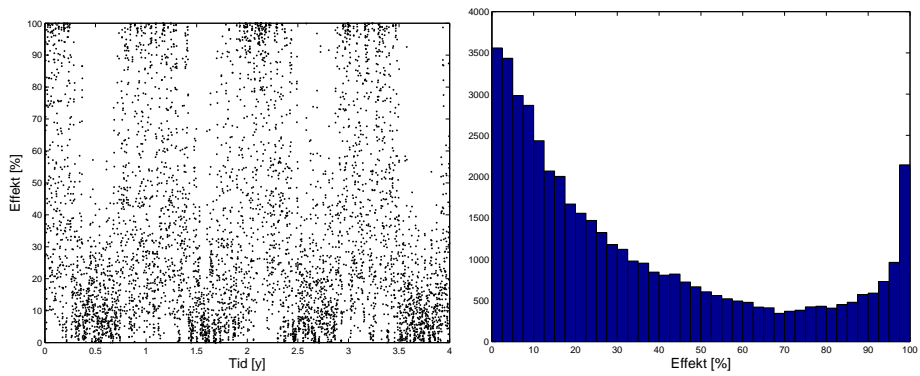
Där $\mu_{\mathbf{p}_i}$ är väntevärdet vid olika tidpunkter, och skattas på samma sätt som i avsnitt 3.3. I figur 3.24 ses resultatet då $K = 45,47$, vilket är medianen av medelenergiproduktionen, och då $K = 36,13$. I samma figur ses även energiproduktionen för den plats som har högst årligt medelvärde.

Effektmedelvärdet blev alltid samma värde som tröskeln K , detta på grund av att högre energi medför högre varians. För $K = 45,47$ blev standardavvikelsen 34,53% och för $K = 36,13$ blev den 30,78%. Som jämförelse kan nämnas att platsen med högst energimedelvärde hade ett medelvärde på 52,69% och en standardavvikelse på 38,46%.

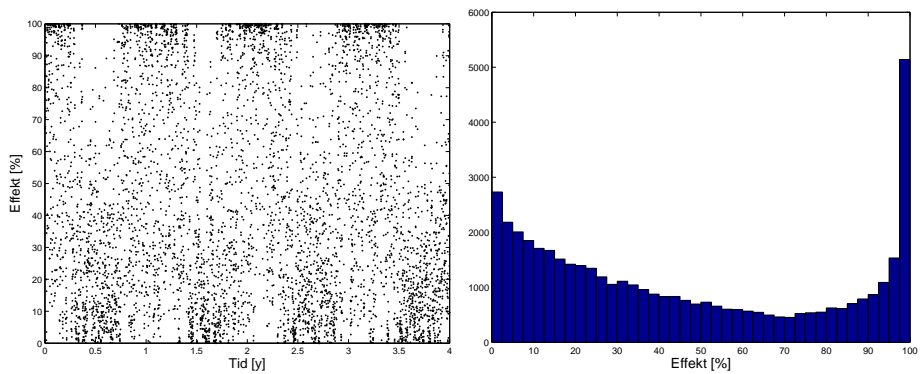
3.5.1 Utvidgning av området

Datamängden erhållen från ECMWF är glest fördelad i rummet. Minimeringen av variansen för den totala energiproduktionen hade med fördel kunnat genomföras på ett större antal platser. Ett sätt att göra detta på är att använda kriging, enligt avsnitt 2.7, för att stokastiskt interpolera mellan givna platser.

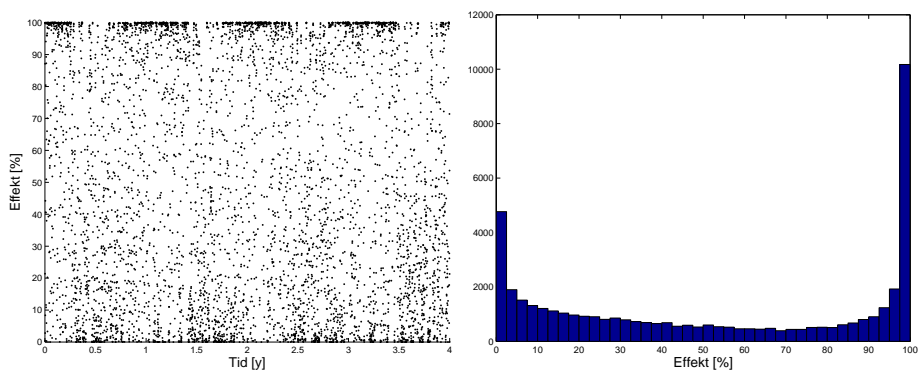
Som underlag för fortsatt arbete har vi konstaterat att områdets semivariogram är anisotropt, se avsnitt 3.2. Vi har även konstaterat att områdets variogram är inhomogent, speciellt för långa avstånd. Däremot är inhomogeniteten mindre för korta avstånd och därmed skulle eventuellt en *moving window*-metod kunna användas. I denna kriging-metod så kan man använda Whittle-modellen då den väl beskriver data i variogrammet för korta avstånd.



(a) Tidsserie för optimal konfiguration då $K=36.13$ (b) Histogram för optimal konfiguration då $K=36.13$.



(c) Tidsserie för optimal konfiguration då $K=45.47$. (d) Histogram för optimal konfiguration då $K=45.47$.



(e) Tidsserie för optimal konfiguration då $K=52.69$ (f) Histogram för optimal konfiguration då $K=52.69$

Figur 3.24: Optimeringsresultat för olika K

4 Diskussion

Resultaten från projektets respektive steg diskuteras här var för sig.

Tillgång till andra datakällor än den från ECMWF (beskriven i avsnitt 3.1) saknades under denna studies utförande. Tanken var att i slutändan testa de utvecklade metoderna på faktiska mätningar. Tyvärr inkom dessa för sent i arbetsprocessen och därför saknades tid för att undersöka dem.

Vårt resultat visar att weibullfördelningen ger en bra beskrivning av vindens årsvisa fördelning, se figur 3.9 och tabell 3.1 på sidan 18. Detta är inte något förvånande, då tidigare studier av vindobservationer även visat detta [5]. Ytterligare en anledning det goda överensstämmandet kan vara att datamängden är simulerad utifrån samma fördelning, men detta är dock inte något vi tittat djupare på.

Weibullfördelningen som beskrivning av vindhastigheter månadsvis har vi däremot inte funnit i några andra källor. Detta antagande ger dock ett bra resultat, med 21 av 365 förkastade hypoteser, se tabell 3.2. Vidare visas i figur 3.14 att detta medför en god överensstämmelse mellan medelvärdesfunktionen skattad med månatliga weibullparameterar och medelvärdet av alla "staplade" år.

En möjlighet vid modelleringen av skalparametern $a(t)$ hade varit att inkludera fler termer i den periodiska funktionen i ekvation (3.8). Denna utvidgning har dock testats, och det resulterade i ett sämre resultat. Först och främst leder det till att fler koefficienter måste skattas, vilket direkt leder till en större osäkerhet. Dessutom fås den oönskvärda effekten att modellen följer det stokastiska beteendet hos tidsserierna. Med få antal parameterar kunde därmed de generella dragen hos $a(t)$ modelleras.

Den ansatta harmoniska funktionen uppfyllde väl sitt syfte, vilket visats i figur 3.16. Detta medförde en god uppskattning av den årliga medelvinden, se tabell 3.3. Anledningen till det goda resultatet med små avvikelser från det riktiga värdet på den årliga medelvinden, är ett starkt beroende mellan den årliga medelvinden och a_0 .

Skattningen av den årliga medeleffekten uppvisar däremot större avvikelser från medeleffekten beräknad direkt på datamängden. Detta kan förklaras med att funktionen för att beräkna effekten i ekvation (2.3) är icke-linjär.

Vidare syns i figur 3.17 och 3.18 att modellens avvikelse från det förväntade värdet ökar österut. Detta skulle kunna förklaras med det faktum att vårt undersökta område är väldigt stort, som längst är det ca 90 mil i öst-västlig riktning.

Att finna parameterar till en modell kräver egentligen bara så många datapunkter som det finns okända parametrar. I vårt fall betraktades a_0 och a_1 som okända, men ändå krävdes ca 100 dagars mätningar innan stabila värden erhöles. Detta är inte krångligare än att datamängden är brusig: man behöver nästan mätdata från en halv periodtid för att kunna särskilja bruset från periodiciteten.

Bayesiansk inferens gav bättre resultat i form av att en kortare tid, knappt 100 dagar krävdes för att skatta parameterarna. Denna metod har flera fördelar då den inte är begränsad på samma sätt som minstakvadrat-metoden, vilken förutsätter att skalparametern $a(t)$ följer den antagna periodiska funktionen. Bayesiansk inferens använder endast detta som expertis och kan därför ge bra resultat även ifall detta antagande inte skulle stämma. Till exempel behöver man inte utnyttja faktorfunktionen i figur 3.20. Istället kan man betrakta denna som ytterligare en parameter, k , i weibullfördelningen $f_X(x|k, a, c)$ i form av en skalning av x till kx . Priorifördelning för k kan då antas vara likformig. Det går även att göra en kvalificerad gissning av den eftersom det återigen endast är en utgångspunkt.

Tillgång till faktiska mätningar saknades, men med en bättre tidsupplösning än mätningar fyra gånger per dag hade de båda metoderna förmodligen gett ett bättre resultat i parameterskattningen. Tidigare studier har lyckats få bra värden på parametrarna i den årliga weibullfördelningen efter endast en månads mättid via bayesiansk parameterskattning [26].

Mätningar med bättre tidsupplösning kommer dock att medföra ett starkare beroende mellan mätningar. För minstakvadrat är detta inte ett problem men i bayesiansk inferens, som kräver oberoende data, kan man vara tvungen att först bearbeta den. I [27] föreslås metoder för hur detta kan genomföras.

Det finns olika sätt att definiera vad som menas med en "optimal" energiproduktion. Vårt sätt att definiera det genom att minimera variansen av den årliga energiproduktionen är endast en av många möjligheter. Två andra möjligheter skulle kunna vara att finna den konfiguration av vindkraftverk som ger den högsta förväntade årliga energiproduktionen eller så stor produktion så möjligt på vintern (då energibehovet är som störst).

Genom att vid optimeringen lägga in en lägsta gräns för den förväntade årliga energiproduktionen (se ekvation (2.7)) utesluts den uppenbara lösningen att lägga alla vindkraftverk på den position som har lägst varians. Detta tillvägagångssätt är mycket allmänt och används bland annat inom portföljoptimering [28].

Då det valda området ligger helt till havs uppvisar vindstyrkan ett liknande beteende i hela området. Att kraftigt reducera fluktuationerna hos energiproduktionen för ett sådant område blir svårt på grund av de positiva korrelationerna mellan platserna. Om det område som optimeras över är tillräckligt utspritt bör det finnas möjliga platser för vindkraftverk som dels är gynnsamma ur effektsynpunkt och dels är negativt korrelerade. Att vårt område saknar denna viktiga egenskap får till följd att väntevärdet för den årliga energiproduktionen blir lägre än vad den annars borde kunna bli. Anledningen till detta är att platser med högt vindmedelvärde oftast även har hög varians, och därför blir energiproduktionen lidande vid minimeringen av dess varians.

På grund av osäkerheten kring datakällan skall inte optimeringen ses som en definitiv konfiguration för minimering av energiproduktionens varians. Resultatet bör istället betraktas ur ett hypotetiskt perspektiv. Antag att man vill ersätta befintliga fossila energikällor med enbart vindkraftverk. Resultat som de i denna studie skulle då kunna ge underlag för att avgöra om det ens är möjligt ur stabilitetssynpunkt. Skulle det visa sig att en optimal, teoretisk konfiguration inte är stabil nog kan man dra slutsatsen att det i praktiken inte är möjligt.

5 Slutsats

Denna studie har visat att weibullfördelningen väl beskriver vindhastighetens årliga och månatliga fördelning. Observationer av vindhastighet har hämtats från en simulerad datamängd tillhandahållen av European Centre for Medium-Range Weather Forecasts (ECMWF). Därmed återstår att undersöka om det samma gäller för faktiska vindobservationer.

Vidare har en tidsberoende modell för weibullfördelningen tagits fram. Denna har verifierats mot beräkningar direkt på datamängden i avseende på årlig medelvind och årlig medeleffekt. Resultatet visade en avvikelse på cirka -1 till +4 % i fallet med medelvind en avvikelse på cirka -5 till +15 % i fallet med medeleffekt.

Studien har även visat att en god skattning av årliga weibullparameterar erhålls både genom en minstakvadrat-anpassning och bayesisk inferens. Efter cirka 100 mätdagar erhålls relativt stabila skattningar av weibullparameterarna.

Trots en optimering av positioner för vindkraftverk i området, erhöles en fluktuerande energiproduktion. Detta förklaras med att vindhastigheterna är för korrelerade.

Referenser

- [1] Azar, C. (2008) Makten över klimatet. Bonnier.
- [2] Statens energimyndighet. (2011) *Vindkraftstatistik*.
<http://energimyndigheten.se/PageFiles/110/Vindkraftsstatistik2011.pdf>.
(Hämtat: 2012-02-04)
- [3] Holmqvist, S och Hedberg, P. (2010) *Åsikter om energi och kärnkraft*
http://www.som.gu.se/digitalAssets/1303/1303503_2010_asikter_om_energi_o_karnkraft.pdf Hämtad 2012-03-15.
- [4] European wind power association. (2011) *Wind in power. 2010 European statistics*
http://ewea.org/fileadmin/ewea_documents/documents/statistics/EWEA_Annual_Statistics_2010.pdf Hämtad 2012-03-15.
- [5] Yilmaz, V. och Celik, H.E. (2008) A statistical approach to estimate the wind speed distribution: the case of Gelibolu region. *Doğuş Üniversitesi Dergisi*, Vol 9, Nummer 1, Sid 122-132
- [6] Holmgren, E. (2009). *Risk indices for the estimation of uncertainty in wind power predictions based on ensembles of numerical weather predictions*. Göteborg: Chalmers tekniska högskola. (Mastersuppsats vid Matematiska Vetenskaper).
- [7] Troen, I., Petersen, L. (1989) *European Wind Atlas*. Laursen: Tønder.
- [8] Andersson, N. (2008) *Stor andel vindkraft ur ett marknads- och teknikperspektiv*.
http://www.elforsk.se/Global/Vindforsk/Rapporter%20fran%20Vindforsk%20II/TR7_08_stor%20andel%20vind.pdf. (Hämtat: 2012-03-02)
- [9] Kiss, P och Janosi, I. M. (2008) Limitations of wind power availability over Europe: a conceptual study. *Nonlinear Processes in Geophysics*. Publicerad 2008-11-03.
<http://www.nonlin-processes-geophys.net/15/803/2008/npg-15-803-2008.pdf>
- [10] Rychlik, I och Rydén, J. (2006) *Probability and Risk Analysis: An Introduction for Engineers*. Springer.
- [11] Rice, J.A. (2006) *Mathematical statistics and data analysis*. Tredje upplagan. Thomson Learning.
- [12] Lee, P.M. (2004) *Bayesian Statistics: An Introduction*. Tredje upplagan. John Wiley & Sons.
- [13] Kstest. *MATLAB R2012a Documentation*.
<http://www.mathworks.se/help/toolbox/stats/kstest.html> (Hämtat: 2012-04-28).
- [14] Devore, J.L. och Farnum, N.R. (1999) *Applied statistics for engineers and scientists*. Duxbury Press Pacific Grove, CA
- [15] Hsu, H. (2010) *Schaum's Outlines: Probability, Random Variables, and Random Processes*, 2:a upplagan. McGraw-Hill.
- [16] Cressie, N. (1991) *Statistics for spatial data*. Wiley-Interscience
- [17] Guttorp, P. och Sampson, P.D. (1994) Methods for estimating heterogeneous spatial covariance functions with environmental applications. *Handbook of Statistics*, vol 12, s 661-689
- [18] Marchant, B.P., Lark, R.M. (2007) The Matérn variogram model: Implications for uncertainty propagation and sampling in geostatistical surveys. *Geoderma*, vol 140, nummer 4, s 337-345.
- [19] Venkatram, A. (1988) On the use of kriging in the spatial analysis of acid precipitation data. *Atmospheric Environment* (1967), vol 22, nummer 9, s 1963-1975.

- [20] Haas, T. (1990) Kriging and automated variogram modeling within a moving window, vol 24, no 7, s 1759 - 1769.
- [21] European Centre for Medium-Range Weather Forecasts (ECMWF). ECMWF Overview. <http://www.ecmwf.int/about/overview/> (Hämtat 2012-04-06).
- [22] Berrisford, P., Dee, D., Poli, Paul et al. (2011) *The ERA-interim archive, version 2.0*. European Centre for Medium Range Weather Forecasts. http://www.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_1_v2.pdf (Hämtat: 2012-04-05)
- [23] Olof Lodén, L. (2012) Ut. *Nationalencyklopedin*. <http://www.ne.se/lang/ut/336951> (Hämtat 2012-04-30)
- [24] European Centre for Medium-Range Weather Forecasts (ECMWF). ERA Project. <http://www.ecmwf.int/about/overview/> (Hämtat 2012-04-06).
- [25] Psi. *MATLAB R2012a Documentation*. <http://www.mathworks.se/help/techdoc/ref/psi.html>. (Hämtat 2012-03-11).
- [26] Erto, P., Lanzotti, A,m Lepore, A. (2010). Wind speed parameter estimation from one-month sample via Bayesian Approach. *Quality and Reliability Engineering International*. Publicerad 2010 på www.wileyonlinelibrary.com DOI: 10.1002/qre.1184. Hämtat (2012-04-23).
- [27] Ramirez, P och Carta, A. (2005) Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. *Energy Conversion and Management*. Vol 46, nummer 15-16, s 2419-2438
- [28] Merton, R. (1972) An Analytic Derivation of the Efficient Portfolio Frontier. *The Journal of Financial and Quantitative Analysis*, vol 7, no 4, s 1851-1872.
- [29] wgrib. National Weather Service: Climate Prediction Center. *wgrib*. <http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html> (Hämtat 2012-02-13)
- [30] World meteorological organization. *A GUIDE TO THE CODE FORM FM 92-IX Ext. GRIB Edition 1*. <http://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html> (Hämtat 2012-02-15)
- [31] Andreasson, N., Evgrafov, A., och Patriksson, M. (2007) *An Introduction to Continuius Optimization*. Studentlitteratur AB.

A Fördelningar

Nedan följer en kort sammanfattning över de fördelningar som använts i rapporten [10].

Weibullfördelning Denna fördelning definieras i ekvation (2.4) på sida 5.

Normalfördelning

Normalfördelningen har en täthetsfunktion enligt

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{A.1})$$

med väntevärde μ och varians σ^2 .

Rayleighfördelning

Rayleighfördelningen är ett specialfall av Weibullfördelningen med formparametern $c = 2$.

Gammafördelning

Gammafördelningen med parametrar α och β definieras utifrån sin täthetsfunktion

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}; \quad 0 \leq x < \infty \quad (\text{A.2})$$

med väntevärde $\alpha\beta$ och varians $\alpha\beta^2$.

Lognormal

En fördelning kallas lognormalfördelad om dess logaritm är normalfördelad. En sådan fördelning har en täthetsfunktion, med parametrar μ och σ , enligt

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \quad (\text{A.3})$$

vilken följer från normalfördelningens täthetsfunktion genom ett variabelbyte.

Birnbaumsaunders

Birnbaum-Saundersfördelningen definieras utifrån täthetsfunktionen

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(\sqrt{x/\beta} - \sqrt{\beta/x})^2}{2\gamma^2}\right) \frac{\sqrt{x/\beta} - \sqrt{\beta/x}}{2\gamma x} \quad (\text{A.4})$$

med skalparameter $\beta > 0$, formparameter $\gamma > 0$ för positiva x .

B Dataomvandling och inläsning i MATLAB

Data som laddades ner från [24] var av formatet `.gribd` och behövde därför konverteras till ett annat format för att kunna läsas in i MATLAB. Ett program för att göra detta är `wgrib`[29].

Vi laddade ner källkoden i C och kompilerade det med kommandot `gcc`. Resultatet blev ett exekverbart program `wgrib` som kunde extrahera data från våra nedladdade `grib`-filer enligt:

```
wgrib -s data.grib | grep ":10U:" | wgrib -i -o output.dat -text data.grib
```

Där den första delen flaggan `-s` i princip endast gör så att innehållet som finns i filen `data.grib` skrivs ut. Nästa *pipe* fångar upp nyckelordet "10U" ur datamängden, detta betyder alla dataposter som beskriver vind som blåser i nord/sydlig-riktning. Den sista *pipen* står för upppackandet och konverteringen av datamängden: resultatet blir en fil i formatet `.dat` där datamängden lagras i text-format.

Denna `.dat`-fil kunde nu läsas in i MATLAB med kommandot `load` eller dylikt. För stora filer blir dock inläsningen väldigt långsam, eftersom att MATLAB måste konvertera om texten till siffror. Ett mer effektivt sätt som senare användes för att konvertera `.grib`-filerna till ett binärt format, och läsa in det i MATLAB med `fread`. För mer information rörande `wgrib`, skriv bara `./wgrib` i terminalen efter kompilering eller läs mer på hemsidan, [29].

C Metropolisalgoritmen

Ett allmänt problem i statistik är att generera slumpstal utifrån en sannolikhetsfördelning $f_X(x)$. Metropolis-Hastingalgoritmen (MH) gör detta genom att diskretisera utfallsrummet för X i sannolikhetsfördelningen och betrakta denna partition som tillstånd i en markovkedja. Genom att välja Markovkedjan enligt MH-algoritmen, kommer denna markovkedja ha samma stationära sannolikhetsfördelning som $f_X(x)$. Algoritmen är enligt [12] följande

1. Välj en godtycklig s.k. proposal transition matrix Q med element $Q_{a,b} = Q(a,b)$. Denna matrix måste vara stokastisk och symmetrisk.
2. välj ett initialt x_0 som ligger i utfallsrummet för kedjorna/täthetsfunktionen
3. for $i=1:n$

 Sampla x från $Q(x_i, x)$. Slumpa sedan ett likförmigt fördelat tal u i intervallet $[0, 1]$ om

$$u < \frac{f_X(x)}{f_X(x_i)} \tag{C.1}$$

 ta $x_{i+1} = x$, annars ta $x_{i+1} = x_i$.

4. end

På detta sätt fås en vektor med tal som är samplad från fördelningen $f_X(x)$. Observera att kvoten i ekvation (C.1) medför att man endast behöver ha tillgång till en funktion som är propotionell mot $f_X(x)$. För en mer teoretisk diskussion om vilka kriterier markovkedjan behöver att uppfylla samt varför algoritmen uppfyller dessa kan ses i [12].

^dEnligt engelska wikipedia-artikeln med samma namn kan `grib` stå för antingen *GRidded Binary* eller *General Regularly-distributed Information in Binary Form*. Formatet har standardiserats av World Meteorological Organization (<http://www.wmo.int/>), och används ofta för att lagra ett stort antal väderobservationer effektivt.[30]

D Kriging, fortsättning

För att härleda ekvationer för λ_i manipuleras först uttrycket i ekvation (2.21). Denna härledning är i stora drag samma som i [19]. En utvidgning av kvadraten i ekvation (2.20) ger

$$\mathbf{E}\left[\underbrace{\sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j X(\mathbf{p}_j) X(\mathbf{p}_i)}_I - 2X(\mathbf{p}_0) \underbrace{\sum_{j=1}^m \lambda_j X(\mathbf{p}_j)}_{II} + \underbrace{X(\mathbf{p}_0)^2}_{III}\right]. \quad (\text{D.1})$$

Termen i kan förenklas enligt

$$\sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j \frac{1}{2} (X(\mathbf{p}_j)^2 + X(\mathbf{p}_i)^2 - (X(\mathbf{p}_j) - X(\mathbf{p}_i))^2). \quad (\text{D.2})$$

Genom att utföra multiplikationen med parentesen och att använda sig av

$$\sum_{i=1}^m \lambda_i = 1 \quad (\text{D.3})$$

och

$$\sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j X(\mathbf{p}_j) = \sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j X(\mathbf{p}_i) \quad (\text{D.4})$$

förenklas uttrycket i till

$$\sum_{i=1}^m \lambda_i X(\mathbf{p}_j)^2 - \frac{1}{2} \sum_{i=1}^m \lambda_i \sum_{i=1}^m \lambda_j (X(\mathbf{p}_j) - X(\mathbf{p}_i))^2. \quad (\text{D.5})$$

Den första termen i denna ekvation kan kvadratkompletteras med ii och iii i ekvation (D.1). Variansen V kan då skrivas som

$$\mathbf{E}\left[-\frac{1}{2} \sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j (X(\mathbf{p}_j) - X(\mathbf{p}_i))^2 + \sum_{i=1}^m \lambda_j (X(\mathbf{p}_j) - X(\mathbf{p}_0))^2\right] = -\sum_{i=1}^m \lambda_i \sum_{j=1}^m \lambda_j \gamma_{ij} - 2 \sum_{j=1}^m \lambda_i \gamma_{i0} \quad (\text{D.6})$$

Den sista likheten fås via definitionen av variogram i ekvation (2.15). För att förenkla beteckningar är $\gamma(\mathbf{p}_i, \mathbf{p}_j) = \gamma_{ij}$. Enligt lagrange multiplikatormetod fås optimala λ genom att lösa följande ekvationsystem. (se följande lärobok i optimering [31], för mera detaljer om hur detta steget genomförs)

$$\begin{aligned} \sum_{j=1}^m \lambda_j \gamma_{ij} + \mu &= \gamma_{i0} \quad i = 1, \dots, m \\ \sum_{j=1}^m \lambda_j &= 1 \end{aligned} \quad (\text{D.7})$$

E Urval av MATLAB-program

E.1 Uppskatta parametrar - kortast tid

```
1 %Författare Fredrik
2
3 clc
4 clear all
5
6 %Inläsning av data från koordinat
7 b = mat_inlasning('HCNorge.mat',7,10);
8
9 %Antal mätpunkter på ett år
10 N = 365*4;
11
12 %Datamängd att använda
13 data = b(1:1*N);
14
15 %dagar som mätts totalt
16 tid = floor(length(data)/4);
17
18 % Används för att uppskatta wbl-param
19 % i ett fönster på 2*L,
20 % steglängd = steg/4 dagar
21 L = 2*4*15;
22 steg = 2*4;
23
24 %förlökering
25 param = zeros(ceil((length(data)-2*L)/steg),2);
26
27 %Skapa wbl-param att senare anpassa mot
28 c = 1;
29 for i = L+1:steg:length(data)-L
30
31     % Data att använda för wbl-anpassning
32     % i ett fönster som är 2*L långt
33     X = data(1,i-L:i+L);
34
35     % Beräkna & Spara resultatet
36     param(c,:) = wblfit(X);
37
38     c = c + 1;
39
40 end
41
42
43 [M, V] = wblstat(param(:,1),param(:,2));
44
45 %Data att anpassa mot med fmincon
46 ydata = param(:,1)';
47
48 % tidsvektor: s skalan på x-axeln blir i dagar
49 xdata = linspace(1,tid,length(ydata));
50 xdata2 = linspace(1,tid,length(param));
51
52 options = optimset('MaxFunEvals',5000, 'TolFun', 1e-10, ...
53     'MaxIter',1200, 'Display', 'off', 'Algorithm','interior-point');
54
55
56 %PARAMETRAR och GRANSER i MODELLEN
57 % amplitud, offset
58 lb = [0, 6]
59 ub = [5, 12]
60
61 T = 365; %Periodtid i dagar
62
63 % startpunkt
64 x0 = [2,8];
65
```

```

66
67 %Var anpassningen skall ske. St rre evalSteg => snabbare
68 evalSteg = 2;
69 evaluate = 1:evalSteg:length(ydata);
70
71
72
73 param_min = zeros(length(evaluate),length(x0));
74
75 c = 1;
76 stop = length(evaluate);
77
78 %2.5697    0.0171   -0.1109   10.2699
79
80 for k = evaluate
81
82     done = c/stop
83
84     % Funktionen att f rs ka anpassa med
85     f = @(x, param) param(1)*cos(2*pi/365.2*x - 0.1109) + param(2);
86
87     % Mlfunktion, att minimera, "minsta kvadrat"
88     % Hur mycket funktionen(param) avviker fr n data
89     res = @(param) norm(f(xdata(1:k),param) - ydata(1:k));
90
91     if k > 1
92         % Anv nder f reg ende resultat som startpunkt
93         x0 = param_min(c-1,:);
94     end
95
96
97     [param_min(c,:), fval(c)] = fmincon(res, x0, [],[],[],[],lb,ub,[],options);
98
99     c = c+1;
100    clc
101 end
102
103
104 %Utvecklingen av a_0 och a_1 ligger nu i param_min

```

E.2 Anpassa modell och verifiera

```

1 %Forfattare Fredrik
2
3 clc
4 clear all
5
6 %loop-variabel, f r att spara ned resultat
7 q = 0;
8
9 % ver  omr dets positioner
10 for i = 1:9
11     for j = 1:13
12
13         q = q+1;
14
15         %Inlasning, b blir en radvektor med tidsserien associerad med i,j
16         b = mat_inlasning('HCNorge.mat',i,j);
17
18         %Antal datapunkter p ett r
19         N = 365*4;
20
21         %Datamngd att anv nda
22         data = b(22*N:23*N);
23
24         %" rlig wbl-f rdelning"
25         konst = wblfit(data);
26
27         valt_c = konst(2);
28

```

```

29     tid = floor(length(data)/4);
30
31     c = 1;
32
33     %Wbl-anpassning till en m nad , med en dag fram t hela tiden
34     stegLangd = 4*1;
35     L = 1*15*4;
36     steg = 4;
37
38     data = [data(:,end-L+1:end) data data(:,1:L)];
39
40     %preallocate
41     param = zeros(ceil((length(data)-2*L)/steg),2);
42     size(param)
43
44     for k = L+1:steg:length(data)-L
45         %k/length(data)
46
47         %Data att anv nda f r wbl-anpassning
48         X = data(k-L:k+L);%A_ex(:,i:i+2*L);
49
50         % Omforma X till radvektor, ty wblfit vill ej ha en matris
51         X =reshape(X, [1 numel(X)]);
52
53         %Spara resultatet
54         param(c,:) = wblfit(X);
55         c = c + 1;
56         clc
57
58     end
59
60     [M,V]= wblstat(param(:,1),param(:,2));
61
62     ydata = param(:,1)';
63     xdata = linspace(1+L/4,tid-L/4,length(param));
64
65     options = optimset('MaxFunEvals',50000, 'TolFun', 1e-10, 'MaxIter',22000);
66
67     funk = @(xdata,z) z(1)*cos(2*pi/365*xdata + z(2)) + z(3);
68
69     x0 = [2.44, -0.187, 10.27];
70
71     [x, resnorm] = lsqcurvefit(@(z,xdata) funk(xdata,z) ,x0,xdata, ydata,[],[], options);
72
73     dagar_forskjut = x(3)/(2*pi)*365.3;
74
75     % x inneh ller nu parameterar till den periodiska modellen
76
77     %%%%%%%%%%%
78     % PLOTTNING
79     if 1 == 0
80         f = figure(1);
81         clf
82         plot(xdata,ydata,'k-', 'LineWidth',1)
83         hold on
84         plot(xdata, funk(xdata,x),'r-', 'LineWidth',2);
85         xlabel('Tid [dagar]', 'fontsize',15)
86         L = legend('Skalparam, a(t)', ['Modell:', num2str(x(1)), ...
87             ' * cos(2*pi/T*t + (', ...
88             num2str(x(2)), ')', ' + ', num2str(x(3))]);
89         set(L, 'fontsize',14)
90     end
91
92     %%%%%%%%%%%
93     %VERIFIERING
94     if 1 == 1
95
96         disp('— VIND —');
97         medelvind_av_data(q) = mean(data);
98         medelvind_ur_modellen(q) = mean(funk(xdata,x)*gamma(1+1/valt_c));
99
100        disp('— EFFEKT —');

```

```

101         wdata = wind2power(data);
102
103         medeleffekt_av_data(q) = mean(wdata);
104
105         for t = 1:length(xdata)
106
107             medel_modell_dt(t) = quadl(@(z) wblpdf(z,funk(xdata(t),x), ...
108                 valt_c).*wind2power(z),0,25);
109
110         end
111
112         medeleffekt_ur_modellen(q) = mean(medel_modell_dt);
113
114     end
115
116     %Spara resultat
117     posx(q) = i;
118     posy(q) = j;
119     a0(q) = x(3);
120     valt_vec(q) = valt_c;
121 end
122
123 end
124
125 %Sammanställ resultat
126 diffvind = (medelvind_av_data-medelvind_ur_modellen)./medelvind_av_data;
127 diffeffekt = (medeleffekt_av_data - medeleffekt_ur_modellen)./medeleffekt_av_data;
128
129 result_vind = [posx' posy' a0' valt_vec' medelvind_ur_modellen' ...
130     medelvind_av_data' diffvind']
131
132 result_effekt = [posx' posy' a0' valt_vec' medeleffekt_ur_modellen' ...
133     medeleffekt_av_data' diffeffekt']

```

E.3 Fördelningstest

```

1 %Forfattare Olivia
2
3 function [kolm]=testdistr(data)
4
5     kolm=zeros(4,2);
6
7     [phat, parmci] = wblfit(data);
8     [h,p,ksstat,cv] = kstest(data, [data' wblcdf(data, phat(1), phat(2))'],0.01);
9     kolm(1,:)=[h p];
10
11
12     phat = gamfit(data/max(data));
13     [h,p,ksstat,cv] = kstest(data, [data' gamcdf(data, phat(1), phat(2))'],0.01);
14     kolm(2,:)=[h p];
15
16
17     phat = raylfit(data);
18     [h,p,ksstat,cv] = kstest(data, [data' raylcdf(data, phat)'],0.01);
19     kolm(4,:)=[h p];
20
21
22 end

```

E.4 Visualisering av optimering

```

1 %Forfattare Amir
2
3 %brus r X-mu
4 t=newShape(brus,9,13,43800); %ger ny form
5 co=covar(t);%ber knar kovarianse ger samma resultat som cov i matlab d medel r ungef r =0.
6 covbrus=co;
7 x=kvadptcov(covbrus,medel,36);%optimerar

```

```

8
9 N=newShape(Norge,9,13,45164);%g r om den till en 2d matris
10
11 for i=1:45164
12     bast(i)=Norge(1,1,i);%g r till vektor
13 end
14
15 figure(1)
16 hold on
17 xaxel=linspace(0,4,4*345*4);
18 plot(xaxel,wind2power(bast(1:length(xaxel))),'.k')
19 figure(2)
20 yax=N'*x;
21 plot(xaxel,wind2power(yax(1:length(xaxel))),'.k')
22 figure(3)
23 hist(wind2power(N'*x),40);
24 figure(4)
25 hist(wind2power(bast),40);
26
27 mean((N'*x))
28 dot(medel,x)
29 var(wind2power(N'*x))
30 mean(wind2power(bast))
31 var(wind2power(bast))

```

E.5 Optimering

```

1 %Forfattare Amir
2
3 function [vikter val]=kvadoptcov(C,medel,K)
4
5
6
7
8 H=C;
9 f=zeros(length(medel),1);
10 A=-medel;%- f r att v nda olikheten
11 b=-K;%-max(medel)-0.1;
12 Aeq=ones(1,length(medel));
13 beq=1;
14 lb=zeros(length(medel),1);%
15 %opts = optimset('Algorithm','interior-point','Display','iter');
16 ub=ones(length(medel),1);%detta r on digt d summan skall bli 1
17 x0=ones(length(medel),1)/length(medel);%gissar uniform
18 [vikter val]= quadprog(H,f,A,b,Aeq,beq,lb,ub,x0);
19
20
21
22 end

```

E.6 Bayesianisk parameteranpassning

```

1 %Forfattare Amir
2
3 m=9.4;
4 star=500;%startid,b rjan av sammaren
5 stop=900;%stoptid,slutet av sammaren
6
7
8 x=faktor(start:stop)'.*reshape(Norge(1,1,start:stop),[1,stop-start+1]);
9 xreal=reshape(Norge(1,1,1:1460),[1,1460]);%ett r
10 [t cil]=wblfit(xreal);%ML ver ett r
11 post=@(para) LHood(para(1),para(2),x)*scaleGivenShape(para(1),para(2),m)*prioriShape(2.8,1.9,para(2));
12
13 delta=0.5;
14 proppdf = @(x,y) unifpdf(y-x,-delta,delta); %proposal pdf i mcmc, relativt godtycklig
15 proprnd = @(x) x + rand*2*delta - delta; %slumpgenerator fr n proppdf
16 nsamples = 26000; %antal punkter man vill ha

```

```

17 para= mhsample([11,2.5],nsamples,'pdf',@(para)post(para),'proprnd',proprnd,'symmetric',1); %metropolis
18 para=para(1000:26000,:);%tar inte med de f rsta v rderna om man gissat d ligt
19 fel=(mean(para)./t-1)*100;%relativa felet

```

E.7 Skattning av variogram

```

1  %Forfattare Amir, Olivia
2
3  clc
4  clearvars -except Norge brus
5  hold off
6  close all
7  load 'Vind.mat', load 'HCNorge.mat'
8  ind=0;
9  for row=1:9
10     for col=1:13
11         ind=ind+1;
12         p1=[63 -13 76 5];
13         latpos=[63:1.5:76];
14         longpos=[-13:1.5:5];
15
16         ref(1)=latpos(row);
17         ref(2)=longpos(col);
18         cc=[1 1 0;1 0 1;0 1 1; 1 0 0;0 1 0;0 0 1;0 0 0];
19         % color string
20
21         %hans avst ndsfkn klarar inte byta jordhalva
22         for i=1:length(latpos)
23             for j=1:length(longpos)
24                 avstand(i,j)=longlatDist(ref(1),ref(2),latpos(i),longpos(j));% h r f s en matris av a
25                 variogram(i,j)=var(brus(row,col,:)-brus(i,j,:));
26             end
27         end
28
29         xdata(numel(avstand)*(ind-1)+1:ind*numel(avstand))=reshape(avstand,[1,numel(avstand)]);
30         ydata(numel(avstand)*(ind-1)+1:ind*numel(avstand))=reshape(variogram,[1,numel(avstand)]);
31
32
33     end
34 end
35 [a,c,n] = variogramfit(xdata,ydata); %minstakvadrat tagen fr n file exchange, file-ID:#25948

```